

# Locally Repairable Codes with Unequal Locality Requirements

Geonu Kim, *Student Member, IEEE*, and Jungwoo Lee, *Senior Member, IEEE*

## Abstract

When a node in a distributed storage system fails, it needs to be promptly repaired to maintain system integrity. While typical erasure codes can provide a significant storage advantage over replication, they suffer from poor repair efficiency. Locally repairable codes (LRCs) tackle this issue by reducing the number of nodes participating in the repair process (locality), with the cost of reduced minimum distance. In this paper, we study the tradeoff between locality and minimum distance of LRCs with local codes that have an arbitrary distance requirement. Unlike existing methods where a common locality requirement is imposed on every node, we allow the locality requirements vary arbitrarily from node to node. Such a property can be an advantage for distributed storage systems with non-homogeneous characteristics. We present Singleton-type distance upper bounds and also provide an optimal code construction with respect to these bounds. In addition, the feasible rate region is characterized by a dimension upper bound that does not depend on the distance. In line with the literature, we first derive bounds based on the notion of locality profile, which refers to the symbol localities specified in a minimum sense. Since the notion of locality profile is less desirable than the locality requirement, which all conventional problem formulations are also based on, we provide locality requirement-based bounds by exhaustively comparing over the relevant locality profiles. Furthermore, and most importantly, we derive bounds with direct expressions in terms of the locality requirement.

## I. INTRODUCTION

### A. Background

The fundamental advantage of storing data in a distributed manner is that the risk of failure can be localized, and a catastrophic loss at once of all the stored data can be avoided. Furthermore, reliability of the system can be improved by using erasure codes on the user data across different storage nodes in a *distributed storage system* (DSS). Although maximum distance separable (MDS) codes provide an optimal storage efficiency for a given amount of reliability, they suffer from poor efficiency during *repair* [1], which is the recovery procedure for failed nodes. Even if the number of failed storage nodes is below the erasure tolerance limit of the codes employed, some

This research was supported in part by the Basic Science Research Program (NRF-2015R1A2A1A15052493) through NRF funded by MSIP, the Technology Innovation Program (10051928) funded by MOTIE, the Bio-Mimetic Robot Research Center funded by DAPA (UD130070ID), INMC, and BK21-plus.

The authors are with the Institute of New Media and Communications, Department of Electrical and Computer Engineering, Seoul National University, Seoul, 08826, Korea (e-mail: bdkim@wspl.snu.ac.kr; junglee@snu.ac.kr).

This paper was presented in part at the 2016 Allerton Conference on Communication, Control and Computing.

(or all) of the failed nodes may have to be promptly repaired to maintain system integrity. For a single node repair, while repetition codes only need access to another single node which is just a replica of the failed node, MDS codes are an opposite extreme in that the number of required helper nodes is as large as the dimension (number of information symbols) of the code. Locally repairable codes (LRCs) try to minimize *locality*, which is the number of nodes that are accessed during repair, for given code parameters such as length, dimension, and minimum distance. The tradeoff between locality and other parameters has been studied extensively since the discovery of the Singleton-type bound in [2].

A natural extension to the conventional locality is the  $(r, \delta)$ -locality [3], [4], where more flexible repair options are provided by generalizing the constraint on the minimum distance of local codes to at least  $\delta$  instead of 2 (single parity checks). Such flexibility is beneficial to modern large-scale DSSs where multiple node failures become more common. For example, in conventional optimal LRCs [2], [5], [6] with locality  $r$ , if another node included in the local repair group of a failed node simultaneously fails, repair from  $r$  nodes is no longer valid, and a large number of nodes have to be accessed to perform ordinary erasure correction. On the other hand,  $(r, \delta)$ -LRCs can still perform repair from  $r$  nodes even if  $\delta - 1$  nodes in a local repair group simultaneously fail.

Recently, there has been interest in the case where locality is specified differently for different nodes [7]–[10]. Such situations may occur, for example, when the underlying storage network is not homogeneous. It would also be beneficial in the scenarios where *hot data* symbols require faster repair or reduced download latency [7]. In [7], [9], relevant Singleton-type bounds have been found and some optimal code constructions are also given, which shows the tightness of the bounds.

## B. Contributions and Organization

In this paper, we study the tradeoff between locality and minimum distance for  $(r, \delta)$ -LRCs, where the locality parameter  $r$  is not necessarily the same for each node. Our main contribution is different from previous work on unequal locality [7], [9] in two ways. First, we extend the results on conventional  $r$ -locality to  $(r, \delta)$ -locality. Specifically, our new Singleton-type bound based on the notion of *locality profile* (Theorem 2) includes the bound in [7] as a special case. Second, and more importantly, we present a bound (Theorem 5) whose expression is directly based on the *locality requirement*, making it more desirable than the straightforward bound (Theorem 3) obtained by maximizing over all locality profiles that comply with the specified locality requirement. Moreover, this bound is shown to be tight in the sense that non-trivial codes (Construction 1) that achieve the equality in the bound exist (Theorem 6). We also characterize the feasible rate region by an upper bound on the code dimension, which does not depend on the minimum distance (Theorem 1, 3, and 4).

The rest of this paper is organized as follows. In Section II, we review some important preliminaries. Section III describes the motivation of our work and provides formal definitions for both the locality profile and the locality requirement. Our Singleton-type bounds based on the locality profile and the locality requirement are respectively provided in Section IV and V, together with the corresponding dimension upper bounds. Section VI shows a code construction scheme that is optimal for the bounds in Section V. In Section VII, we provide some further results,

including the optimality of the code construction in Section VI in terms of the bound in Section IV, and a further tightened bound based on the locality requirements with two unequal parameters under a special condition. Finally, the concluding remarks are drawn in Section VIII.

## II. PRELIMINARIES

### A. Notation

We use the following notation.

- 1) For an integer  $i$ ,  $[i] = \{1, \dots, i\}$ .
- 2) A vector of length  $n$  is denoted by  $\mathbf{v} = (v_1, \dots, v_n)$ .
- 3) A matrix of size  $k \times n$  is denoted by  $G = (g_{i,j})_{i \in [k], j \in [n]}$ .
- 4) For sets  $\mathcal{A}$  and  $\mathcal{B}$ ,  $\mathcal{A} \sqcup \mathcal{B}$  denotes the disjoint union, i.e.,  $\mathcal{A} \cup \mathcal{B}$  with further implication that  $\mathcal{A} \cap \mathcal{B} = \emptyset$ .
- 5) For a symbol index set  $\mathcal{T} \subset [n]$  of a code  $\mathcal{C}$  of length  $n$ ,  $\mathcal{C}|_{\mathcal{T}}$  denotes the punctured code with support  $\mathcal{T}$ , and  $G|_{\mathcal{T}}$  is the corresponding generator matrix. Furthermore, we define  $\text{rank}_G(\mathcal{T}) = \text{rank}(G|_{\mathcal{T}})$ .
- 6) For a symbol index set  $\mathcal{T} \subset [n]$  of a linear  $[n, k]$  code  $\mathcal{C}$  constructed via polynomial evaluation on an extension field  $\mathbb{F}_{q^t}$ ,  $\text{rank}_E(\mathcal{T})$  denotes the rank of the evaluation points corresponding to  $\mathcal{C}|_{\mathcal{T}}$  over the base field  $\mathbb{F}_q$ .

### B. Minimum Distance

The minimum distance of linear codes is well known to be characterized by the following lemma [4, Lem. A.1], which is the basis of our minimum distance bounds.

**Lemma 1.** *For a symbol index set  $\mathcal{T} \subset [n]$  of a linear  $[n, k, d]$  code such that  $\text{rank}_G(\mathcal{T}) \leq k - 1$ , we have*

$$d \leq n - |\mathcal{T}|,$$

*with equality if  $\mathcal{T}$  is of largest cardinality.*

Below, we state a lemma (see also [11]) based on Lemma 1 that turns out to be more useful. Note that Lemma 2 can not be derived by simply substituting  $|\mathcal{T}|$  into Lemma 1.

**Lemma 2.** *For a symbol index set  $\mathcal{T} \subset [n]$  of a linear  $[n, k, d]$  code such that  $\text{rank}_G(\mathcal{T}) \leq k - 1$ , let  $\gamma$  be the number of redundant symbols indexed by  $\mathcal{T}$ , i.e.,  $\gamma = |\mathcal{T}| - \text{rank}_G(\mathcal{T})$ . We have*

$$d \leq n - k + 1 - \gamma.$$

*Proof:* Clearly, the set  $\mathcal{T}$  can be enlarged to a set  $\mathcal{T}'$  such that  $\text{rank}_G(\mathcal{T}') = k - 1$ . Make another set  $\mathcal{T}''$  by removing  $\gamma$  redundant symbols from  $\mathcal{T}'$ . Note that  $|\mathcal{T}''| \geq k - 1$  since  $\text{rank}_G(\mathcal{T}'') = k - 1$ . By applying Lemma 1 to the set  $\mathcal{T}'$ , we have

$$\begin{aligned} d &\leq n - |\mathcal{T}'| = n - |\mathcal{T}''| - \gamma \\ &\leq n - k + 1 - \gamma. \end{aligned}$$

■

As an immediate corollary to Lemma 1, we also get the following lemma, which is used when showing the optimal distance property of our code construction.

**Lemma 3.** *For linear  $[n, k, d]$  codes, if  $\text{rank}_G(\mathcal{T}) = k$  for every symbol index set  $\mathcal{T} \subset [n]$  such that  $|\mathcal{T}| = \tau$ , we have*

$$d \geq n - \tau + 1.$$

**Remark 1.** *In Lemma 1, 2, and 3, erasure correction is possible from  $\mathcal{T}$  if and only if  $\text{rank}_G(\mathcal{T}) = k$ . Equivalently, erasure correction is not possible from  $\mathcal{T}$  if and only if  $\text{rank}_G(\mathcal{T}) \leq k - 1$ .*

### C. $(r, \delta)$ -Locality<sup>1</sup>

A linear  $[n, k, d]$  code  $\mathcal{C}$  is said to have *locality  $r$*  (or  $r$ -locality) if every symbol of  $\mathcal{C}$  can be recovered with a linear combination of at most  $r$  other symbols [2]. An equivalent description is that for each symbol index  $i \in [n]$ , there exists a punctured code of  $\mathcal{C}$  with support containing  $i$ , length of at most  $r + 1$  and distance of at least 2. We call such codes  $r$ -LRCs. It has been shown in [2] that the minimum Hamming distance  $d$  of an  $[n, k, d]$   $r$ -LRC is upper bounded by

$$d \leq n - k + 2 - \left\lceil \frac{k}{r} \right\rceil,$$

which reduces to the well-known *Singleton bound* if  $r \geq k$ . Various optimal code constructions achieving the equality in the minimum distance bound have been reported in the literature [2], [5], [6], [12]–[17].

The notion of  $r$ -locality can be naturally extended to  $(r, \delta)$ -locality [3] to address the situation with multiple (local) node failures.

**Definition 1** ( $(r, \delta)$ -locality). *For a symbol with index  $i \in [n]$  of a linear  $[n, k]$  code  $\mathcal{C}$ , suppose there exists a punctured code of  $\mathcal{C}$  with support containing  $i$ , length of  $r + \delta - 1$  and distance of  $\delta$ , i.e., there exists a symbol index set  $\mathcal{S}_i \subset [n]$  such that*

- $i \in \mathcal{S}_i$ ,
- $|\mathcal{S}_i| = r + \delta - 1$ ,
- $d(\mathcal{C}|_{\mathcal{S}_i}) = \delta$ .

*If  $r$  is minimum, the  $i$ th symbol is said to be of  $(r, \delta)$ -locality and we denote it by  $\text{loc}_\delta(i) = r$ . Furthermore, we define*

$$\text{loc}_\delta(\mathcal{C}) = \max_{i \in [n]} \{\text{loc}_\delta(i)\},$$

*and also denote  $\mathcal{C}$  to be of  $(\text{loc}_\delta(\mathcal{C}), \delta)$ -locality.*

<sup>1</sup> To prevent ambiguity, we make our notation and definitions slightly different from that in [3]. However, the expression of *having* locality  $(r, \delta)$  in [3] can be verified to be equivalent to the condition  $\text{loc}_\delta(\cdot) \leq r$  in our notation, for both symbols and codes. In our paper, we use the expression of *satisfying* an  $(r, \delta)$ -locality *requirement* to denote that condition.

For  $2 \leq \delta \leq d$ ,  $\text{loc}_\delta(i)$  in Definition 1 can be shown to be *well-defined*. In particular, considering the process of repeatedly puncturing  $\mathcal{C}$  at a symbol not indexed by  $i$ , each time the minimum distance decreases by at most 1 and ends up with 1. Therefore, we can always find a set  $\mathcal{S}_i$  such that  $d(\mathcal{C}|_{\mathcal{S}_i}) = \delta$ . Further choosing  $\mathcal{S}_i$  to be of minimum cardinality, we have  $\text{loc}_\delta(i) = |\mathcal{S}_i| - \delta + 1$ .

**Remark 2.** By applying the Singleton bound to  $\mathcal{C}|_{\mathcal{S}_i}$  in Definition 1, we get  $\text{rank}_G(\mathcal{S}_i) \leq r$ .

If  $\text{loc}_\delta(\mathcal{C}) \leq r$ , we say that  $\mathcal{C}$  satisfies the  $(r, \delta)$ -locality requirement, and also call it an  $(r, \delta)$ -LRC. It is shown in [3], [4] that the minimum distance of an  $(r, \delta)$ -LRC is upper bounded by

$$d \leq n - k + 1 - \left( \left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1). \quad (1)$$

There are also several optimal code constructions in the literature [3], [4], [6], [12]–[14], [18]–[20] that achieve the equality in (1).

#### D. $r$ -LRC with Unequal Locality Profile

Conventionally, the locality of a code is characterized by a single parameter  $r$ , i.e.,  $r$ -LRC. Although an  $[n, k]$   $r$ -LRC<sup>2</sup> can have different  $\text{loc}_2(i)$  for different  $i \in [n]$ , as long as  $\text{loc}_2(i) \leq r$ , optimal codes generally require  $\text{loc}_2(i) = r$  for every symbol due to the general tradeoff between locality and distance.

Inspired by the notion of *unequal error protection*, [7] considers linear codes where different subsets of symbols have different localities. This feature can be captured by the notion of *locality profile*. In particular, the locality profile of a linear  $[n, k, d]$  code  $\mathcal{C}$  is defined as the vector  $\mathbf{r} = (r_1, \dots, r_n)$ , where  $r_i = \text{loc}_2(i)$ ,  $i \in [n]$ . The locality profile can also be specified as another vector  $\mathbf{n} = (n_1, \dots, n_{r^*})$ , where  $r^* = \max\{r_1, \dots, r_n\}$  and  $n_j$  is the number of symbols such that  $\text{loc}(\cdot)$  equals  $j$ , for  $j \in [r^*]$ . The minimum Hamming distance of  $\mathcal{C}$  is shown to be upper bounded by

$$d \leq n - k + 2 - \sum_{j=1}^{r-1} \left\lceil \frac{n_j}{j+1} \right\rceil - \left\lceil \frac{k - \sum_{j=1}^{r-1} (n_j - \left\lceil \frac{n_j}{j+1} \right\rceil)}{r} \right\rceil, \quad (2)$$

where

$$r = \min \left\{ j \in [r^*] \mid \sum_{j'=1}^j (n_{j'} - \left\lceil \frac{n_{j'}}{j'+1} \right\rceil) \geq k \right\}.$$

Furthermore, this bound is demonstrated to be tight by some optimal code constructions achieving the equality in the bound for some parameter regime.

The work in [9] has investigated the same problem, resulting in a bound with an expression similar to (2), where the difference comes mostly from notational difference. However, after aligning the notation, one can verify that (2) is tighter in general.

<sup>2</sup> $r$ -LRCs can be shown to be a special case of  $(r, \delta)$ -LRCs with  $\delta = 2$ .

### E. Gabidulin Codes

Our optimal code construction is an extension of the LRC construction based on Gabidulin codes [7], [13]. We thus give a brief introduction on Gabidulin codes, including some relevant properties.

Due to the vector space structure of extension fields, an element in  $\mathbb{F}_{q^t}$  can be equivalently expressed as a vector of length  $t$  over the base field  $\mathbb{F}_q$ , i.e.,  $\mathbb{F}_q^t$ . Consequently, a vector  $\mathbf{v} \in \mathbb{F}_{q^t}^n$  can be represented as a matrix  $V \in \mathbb{F}_q^{t \times n}$  where each column vector of the matrix  $V$  corresponds to the vector representation of an element in vector  $\mathbf{v}$ . The rank of the vector  $\mathbf{v}$  is defined as  $\text{rank}(\mathbf{v}) = \text{rank}(V)$ . Furthermore, a *metric* called *rank distance* can be defined for two vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{F}_{q^t}^n$  as

$$d_R(\mathbf{u}, \mathbf{v}) \triangleq \text{rank}(\mathbf{u} - \mathbf{v}) = \text{rank}(U - V).$$

It is easy to see that the rank distance is upper bounded by the Hamming distance, i.e.,  $d_R(\mathbf{u}, \mathbf{v}) \leq d_H(\mathbf{u}, \mathbf{v})$ . Therefore the minimum rank distance of a linear  $[n, k]_{q^t}$  code is also upper bounded by the Singleton bound, and the codes achieving this bound are called maximum rank distance (MRD) codes. Clearly, MRD codes also have the MDS property.

Gabidulin codes [21] are an important class of codes with the MRD property. Similar to Reed-Solomon and other algebraic codes, Gabidulin codes are constructed via polynomial evaluation. However, both the data polynomials and the evaluation points are different. In particular, an  $[n, k, d]_{q^t}$  Gabidulin code ( $t \geq n$ ) is constructed by encoding a message vector  $\mathbf{a} = (a_1, \dots, a_k) \in \mathbb{F}_{q^t}^k$  according to the following two steps.

- 1) Construct a data polynomial  $f(x) = \sum_{i=1}^k a_i x^{q^{i-1}}$ .
- 2) Obtain a codeword by evaluating  $f(x)$  at  $n$  points  $\{x_1, \dots, x_n\} \subset \mathbb{F}_{q^t}$  (or  $\mathbb{F}_q^t$ ) that are linearly independent over  $\mathbb{F}_q$ , i.e.,  $\mathbf{c} = (f(x_1), \dots, f(x_n)) \in \mathbb{F}_{q^t}^n$  with  $\text{rank}(\{x_1, \dots, x_n\}) = n$ .

The data polynomial  $f(x)$  belongs to a special class of polynomials called *linearized polynomials* [22]. The evaluation of a linearized polynomial over  $\mathbb{F}_{q^t}$  is an  $\mathbb{F}_q$ -linear transformation. In other words, for any  $a, b \in \mathbb{F}_q$  and  $x, y \in \mathbb{F}_{q^t}$ , the following holds.

$$f(ax + by) = af(x) + bf(y). \quad (3)$$

The rank distance of any Gabidulin codeword  $\mathbf{c}$  can be shown to meet the Singleton bound by noting that

$$\text{rank}(\mathbf{c}) = \dim(\text{span}(\{f(x_1), \dots, f(x_n)\})) \stackrel{(a)}{=} \dim(f(\text{span}(\{x_1, \dots, x_n\}))) \stackrel{(a)}{\geq} n - (k - 1),$$

where (a) is due to the *rank-nullity theorem* and the fact that the nullity of  $f(\cdot)$  is at most the  $q$ -degree of  $f(\cdot)$ , i.e.,  $k - 1$ .

Although it is sufficient to claim Gabidulin codes to be MDS by their MRD property, a more insightful derivation can be obtained by showing the MDS property directly with the analysis of its erasure correction capability. Specifically, the polynomial  $f(\cdot)$ , and therefore the underlying message vector  $\mathbf{a}$ , can be recovered from evaluations on any  $k$  points  $\{f(y_1), \dots, f(y_k)\}$  that are linearly independent (over  $\mathbb{F}_q$ ), i.e.,  $\text{rank}(\{y_1, \dots, y_k\}) = k$ . This argument is true since the use of the  $\mathbb{F}_q$ -linearity in (3) makes it possible to obtain evaluations at  $q^k$  different points,

from which the polynomial  $f(\cdot)$  of degree  $q^{k-1}$  can be interpolated. Therefore, erasure correction is possible from arbitrary  $k$  symbols of the codeword.

More importantly, note that the evaluation points may differ from the original ones used in the codeword construction. This turns out to be the case for our optimal code construction, where we apply MDS encoding on chunks of a Gabidulin codeword to equip the code with the desired locality property. To analyze the possibility of erasure correction (or decodability) of an erasure pattern of the code, all we need to do is to figure out whether the *remaining rank*, which refers to the rank of the evaluation points corresponding to the remaining symbols of the erasure pattern, is at least  $k$ . Moreover, the term *rank erasure* will refer to the vanishing rank by symbol erasures. The following lemma, which is a special case of [23, Lem. 9], will be used several times in analyzing the distance of our optimal code construction.

**Lemma 4.** *For a vector  $\mathbf{u}$  of length  $k$  with elements being evaluations of a linearized polynomial  $f(\cdot)$  over  $\mathbb{F}_{q^t}$ , such that the evaluation points are linearly independent over  $\mathbb{F}_q$ , let  $\mathbf{v}$  be the vector obtained by encoding  $\mathbf{u}$  with an  $[n, k]_q$  MDS code. Then any  $s$  symbols of the codeword  $\mathbf{v}$  correspond to the evaluations of  $f(\cdot)$  at  $s$  points lying in the subspace spanned by the original  $k$  evaluation points (of  $\mathbf{u}$ ) with rank  $\min(s, k)$ , i.e., for an arbitrary set  $\mathcal{T} \subset [n]$  such that  $|\mathcal{T}| = s$ , we have*

$$\text{rank}_{\mathbb{E}}(\mathcal{T}) = \min(s, k).$$

*Proof:* See Appendix A. ■

### III. $(r, \delta)$ -LRC WITH UNEQUAL LOCALITY

Consider DSSs that have different  $(r, \delta)$ -locality requirements (with fixed  $\delta$ ) for different nodes. In other words, the upper limit for  $\text{loc}_{\delta}(\cdot)$  varies from symbol to symbol for the code employed in the DSS. Among the different locality requirements, let us denote the minimum locality requirement by  $r_{\min}$ . We can employ codes satisfying the  $(r_{\min}, \delta)$ -locality requirement, i.e.,  $(r_{\min}, \delta)$ -LRCs. However, this clearly tends to be an *over-design*, and one may expect improved minimum distance by taking advantage of the looser locality requirements.

To be more specific, suppose for a linear  $[n, k, d]$  code and symbol index sets  $\mathcal{N}_1, \mathcal{N}_2 \subsetneq [n]$  such that  $\mathcal{N}_1 \sqcup \mathcal{N}_2 = [n]$ , we require  $\text{loc}_{\delta}(i_1) \leq r_1$  and  $\text{loc}_{\delta}(i_2) \leq r_2$ , where  $i_1 \in \mathcal{N}_1$ ,  $i_2 \in \mathcal{N}_2$  and  $r_1 < r_2$ . Clearly a code  $\mathcal{C}_1$  such that  $\text{loc}_{\delta}(\mathcal{C}_1) \leq r_1$  satisfies this requirement and by optimal code constructions with respect to (1), we may achieve

$$d = d_1 \triangleq n - k + 1 - \left( \left\lceil \frac{k}{r_1} \right\rceil - 1 \right) (\delta - 1).$$

On the other hand, any code fulfilling the different locality requirements also clearly satisfies the  $(r_2, \delta)$ -locality requirement, and we therefore have, again by (1),

$$d \leq d_2 \triangleq n - k + 1 - \left( \left\lceil \frac{k}{r_2} \right\rceil - 1 \right) (\delta - 1).$$

Note that  $d_1 \leq d_2$  where equality does not hold in general. Now the question is as follows: can we construct codes achieving a distance larger than  $d_1$ , and at the same time, do we have a distance bound tighter than  $d_2$ ? To provide

an answer to the question, we first introduce the notion of *locality profile* and *locality requirement* for linear  $[n, k]$  codes. Note that some redundant parameters are further defined, which makes various expressions more compact.

**Definition 2** (Locality profile). *Given a linear  $[n, k]$  code  $\mathcal{C}$ , let  $r_i$  be such that  $\text{loc}_\delta(i) = r_i$ ,  $i \in [n]$ . Denoting  $\mathcal{N}_j = \{i \in [n] \mid \text{loc}_\delta(i) = j\}$ ,  $j \in [r^*]$ , where  $r^* = \max\{r_1, \dots, r_n\}$ , the locality profile of  $\mathcal{C}$  is defined as an ordered pair  $(\mathbf{n}, \delta)$  where  $\mathbf{n} = (n_1, \dots, n_{r^*})$  such that  $n_j = |\mathcal{N}_j|$ . Furthermore, define*

- integers  $p_j, q_j$  such that  $n_j = p_j(j + \delta - 1) + q_j$  and  $0 \leq q_j \leq j + \delta - 2$ ,
- $m_j \triangleq \frac{n_j}{j + \delta - 1} = p_j + \frac{q_j}{j + \delta - 1}$ ,
- $k_j \triangleq \begin{cases} \lfloor m_j \rfloor j & \text{if } 0 \leq q_j \leq \delta - 2, \\ n_j - \lceil m_j \rceil (\delta - 1) & \text{if } \delta - 1 \leq q_j \leq j + \delta - 2. \end{cases}$

**Definition 3** (Locality requirement). *The locality requirement is defined as an ordered pair  $(\mathbf{n}, \delta)$  where  $\mathbf{n} = (n_1, \dots, n_{r^*})$ , such that  $n_{r^*} \neq 0$ . A linear  $[n, k]$  code  $\mathcal{C}$  is said to satisfy the locality requirement  $(\mathbf{n}, \delta)$ , if there exist symbol index sets  $\mathcal{N}_j \subset [n]$  of  $\mathcal{C}$  such that  $|\mathcal{N}_j| = n_j$ ,  $n = \sum_{j=1}^{r^*} n_j$ ,  $\mathcal{N}_j \cap \mathcal{N}_{j'} = \emptyset$  if  $j \neq j'$ , and each symbol with index  $i \in \mathcal{N}_j$  satisfies the  $(j, \delta)$ -locality requirement, i.e.,  $\text{loc}_\delta(i) \leq j$ ,  $j, j' \in [r^*]$ . Furthermore, as for the locality profile (Definition 2), parameters  $p_j, q_j$  and  $m_j$  are implicitly defined for a locality requirement  $(\mathbf{n}, \delta)$ .*

In our study, a direct analysis of the distance characteristics of LRCs based on a given locality requirement appears to be infeasible due to its inherent ambiguity. In particular, a symbol satisfying the  $(r_1, \delta)$ -locality requirement also satisfies, by definition, the  $(r_2, \delta)$ -locality requirement. Note that to tighten the straightforward minimum distance upper bound given by  $d_2$ , it becomes essential to treat symbols with different locality requirements separately. In general, a symbol identified as satisfying the  $(r_2, \delta)$ -locality requirement can participate in the local repair of other symbols satisfying the  $(r_1, \delta)$ -locality requirement, which means it also satisfies the  $(r_1, \delta)$ -locality requirement. This fact makes it very difficult to separate symbols identified by the  $(r_2, \delta)$ -locality requirement from symbols with the  $(r_1, \delta)$ -locality requirement. To handle this issue, we first assume the knowledge of  $\text{loc}_\delta(\cdot)$  for every symbol, i.e., the locality profile, and later proceed to the original problem formulation based on the locality requirement. The key characteristic of the locality profile that makes the problem more tractable is given in the following remark.

**Remark 3.** *Suppose that the  $i$ th symbol is of  $(r, \delta)$ -locality, i.e.,  $\text{loc}_\delta(i) = r$ . Clearly, this symbol cannot participate in the repair process of the  $j$ th symbol of  $(\tilde{r}, \delta)$ -locality such that  $\tilde{r} < r$ , since we otherwise have a contradiction such that  $\text{loc}_\delta(i) \leq \tilde{r} < r$ . Therefore, the symbols indexed by  $\mathcal{S}_j$  (see Definition 1) are guaranteed not to be of  $(r, \delta)$ -locality.*

Another view of our approach is to first focus on codes that meet the locality requirement with equality. Therefore, an analysis of the minimum distance based on the locality profile not only makes the problem more tractable, but is also reasonable because larger locality tends to provide larger minimum distance. However, note that a more useful problem setting is based on the locality requirement, which also corresponds to the conventional problem formulations with equal locality requirements [2], [3].



While a given code has a unique locality profile, it satisfies (infinitely) many locality requirements, and more importantly, codes of different locality profiles can satisfy a given locality requirement. Furthermore, it is clear that there always exists a code satisfying a locality requirement, but this may not be the case for a given locality profile.

**Example 1.** Consider a locality requirement  $(\mathbf{n}, \delta)$  where  $\mathbf{n} = (n_1, n_2, n_3) = (0, 1, 1)$ . For a linear code  $\mathcal{C}$  of length  $n = n_1 + n_2 + n_3 = 2$  such that  $\text{loc}_\delta(\mathcal{C}) \leq 3$ , valid values for the parameter  $\hat{\mathbf{n}}$  of its locality profile  $(\hat{\mathbf{n}}, \delta)$  are  $(2, 0, 0)$ ,  $(1, 1, 0)$ ,  $(1, 0, 1)$ ,  $(0, 2, 0)$ ,  $(0, 1, 1)$  and  $(0, 0, 2)$ , of which all except the last one satisfy the locality requirement  $(\mathbf{n}, \delta)$ .

#### IV. UPPER BOUNDS BASED ON LOCALITY PROFILE

In this section, we provide a minimum distance upper bound for  $(r, \delta)$ -LRCs based on their *locality profile*. The derivation starts by defining some rank-related auxiliary parameters and proceeds in two steps. First, a bound based on the auxiliary parameters is established, and the auxiliary parameters are subsequently eliminated in the bound expression by utilizing an upper bound on themselves. As a by-product, we also give a dimension upper bound which does not depend on the minimum distance.

Given Definition 2, the auxiliary parameters are defined as follows. Note that, these auxiliary parameters are always assumed to be implicitly defined whenever we specify a locality profile  $(\mathbf{n}, \delta)$ .

**Definition 4** (Auxiliary parameters). *Given a locality profile  $(\mathbf{n}, \delta)$ , define the following rank-related parameters, for  $j \in [r^*]$ .*

$$\xi_j \triangleq \text{rank}_G\left(\bigsqcup_{j'=1}^j \mathcal{N}_{j'}\right) - \text{rank}_G\left(\bigsqcup_{j'=1}^{j-1} \mathcal{N}_{j'}\right).$$

The following lemma gives a distance upper bound with the auxiliary parameters appearing in the expression. The proof is built on the algorithmic technique originally proposed in [2], which has been widely used in the literature [4], [7], [9] with some modification.

**Lemma 5.** *The minimum Hamming distance of linear  $[n, k, d]$  codes with locality profile  $(\mathbf{n}, \delta)$  is upper bounded by*

$$d \leq n - k + 1 - \sum_{j=1}^{\rho-1} (n_j - \xi_j) - \left( \left\lceil \frac{k - \sum_{j=1}^{\rho-1} \xi_j}{\rho} \right\rceil - 1 \right) (\delta - 1),$$

where

$$\rho = \min\{j' \in [r^*] \mid \sum_{l=1}^{j'} \xi_l = k\}.$$

*Proof:* By using Algorithm 1, we build a set  $\mathcal{T} \subset [n]$  such that  $\text{rank}_G(\mathcal{T}) \leq k-1$ , and apply Lemma 2 to obtain the distance upper bound. In the algorithm,  $\mathcal{S}_i$  denotes the support of the punctured code by which  $\text{loc}_\delta(i) = j$ . Note that  $\mathcal{S}_i \subset \bigsqcup_{j'=1}^j \mathcal{N}_{j'}$  due to Remark 3. Therefore,  $\mathcal{Q}_l \subset \bigsqcup_{j'=1}^j \mathcal{N}_{j'}$  and  $\text{rank}_G(\mathcal{Q}_l) \leq \text{rank}_G(\bigsqcup_{j'=1}^j \mathcal{N}_{j'})$ . The condition in Step 2 ensures that it is always possible to pick a suitable  $i$  in Step 3. The algorithm iterates until

---

**Algorithm 1** Used in the Proof of Lemma 5 and Lemma 6
 

---

```

1: Let  $\mathcal{Q}_0 = \bigsqcup_{j'=1}^{j-1} \mathcal{N}_{j'}$ ,  $l = 0$ 
2: while  $\text{rank}_G(\mathcal{Q}_l) < \text{rank}_G(\bigsqcup_{j'=1}^j \mathcal{N}_{j'})$  do
3:   Pick any  $i \in \mathcal{N}_j \setminus \mathcal{Q}_l$  such that  $\text{rank}_G(\mathcal{Q}_l \cup \mathcal{S}_i) > \text{rank}_G(\mathcal{Q}_l)$ 
4:    $l = l + 1$ 
5:    $\mathcal{Q}_l = \mathcal{Q}_{l-1} \cup \mathcal{S}_i$ 
6: end while
7:  $L = l$ 

```

---

$l = L$ , where  $\text{rank}_G(\mathcal{Q}_L) = \text{rank}_G(\bigsqcup_{j'=1}^j \mathcal{N}_{j'})$ . Note that we have

$$L \geq \left\lceil \frac{\text{rank}_G(\mathcal{Q}_L) - \text{rank}_G(\mathcal{Q}_0)}{j} \right\rceil = \left\lceil \frac{\xi_j}{j} \right\rceil, \quad (4)$$

since for the incremental rank in each iteration, we have

$$\text{rank}_G(\mathcal{Q}_l) - \text{rank}_G(\mathcal{Q}_{l-1}) \leq \text{rank}_G(\mathcal{S}_i) \stackrel{(a)}{\leq} j,$$

$l \in [L]$ , where (a) is due to Remark 2.

We claim that

$$|\mathcal{Q}_l| - |\mathcal{Q}_{l-1}| \geq \text{rank}_G(\mathcal{Q}_l) - \text{rank}_G(\mathcal{Q}_{l-1}) + \delta - 1. \quad (5)$$

To see why (5) holds, first note that in the context of the punctured code corresponding to  $\mathcal{S}_i$ , the symbols indexed by an arbitrary subset of  $\mathcal{S}_i$  with the size of  $\delta - 1$  are redundant since  $d(\mathcal{C}|_{\mathcal{S}_i}) = \delta$ . We have  $|\mathcal{Q}_l| - |\mathcal{Q}_{l-1}| = |\mathcal{Q}_l \setminus \mathcal{Q}_{l-1}| \geq \delta$ , since otherwise we must have  $\text{rank}_G(\mathcal{Q}_l) = \text{rank}_G(\mathcal{Q}_{l-1})$ , due to the fact that  $\mathcal{Q}_l \setminus \mathcal{Q}_{l-1} \subset \mathcal{S}_i$ . This is contradictory to the condition in Step 3. Now, out of the  $|\mathcal{Q}_l| - |\mathcal{Q}_{l-1}| \geq \delta$  incremental symbols in the set  $\mathcal{Q}_l$ , at least  $\delta - 1$  symbols are redundant since they are already redundant in the context of  $\mathcal{S}_i \subset \mathcal{Q}_l$ . Therefore,  $\text{rank}_G(\mathcal{Q}_l) - \text{rank}_G(\mathcal{Q}_{l-1}) \leq |\mathcal{Q}_l| - |\mathcal{Q}_{l-1}| - (\delta - 1)$ .

Next, set  $j = \rho$  in the algorithm. By (4) and the definition of  $\rho$ , we have

$$L \geq \left\lceil \frac{\xi_\rho}{\rho} \right\rceil = \left\lceil \frac{k - \sum_{j=1}^{\rho-1} \xi_j}{\rho} \right\rceil.$$

Let  $\mathcal{T} = \mathcal{Q}_L$  where

$$l = \left\lceil \frac{k - \sum_{j=1}^{\rho-1} \xi_j}{\rho} \right\rceil - 1.$$

Since  $l \leq L - 1$ , we have

$$\begin{aligned} \text{rank}_G(\mathcal{T}) &\leq \text{rank}_G\left(\bigsqcup_{j=1}^{\rho} \mathcal{N}_j\right) - 1 = \sum_{j=1}^{\rho} \xi_j - 1 \\ &= k - 1. \end{aligned}$$

We conclude the proof by noting that the number of redundant symbols indexed by  $\mathcal{T}$  is

$$\begin{aligned}
\gamma &= |\mathcal{T}| - \text{rank}_G(\mathcal{T}) = |\mathcal{Q}_l| - \text{rank}_G(\mathcal{Q}_l) \\
&= \sum_{l'=1}^l (|\mathcal{Q}_{l'}| - |\mathcal{Q}_{l'-1}|) + |\mathcal{Q}_0| - \sum_{l'=1}^l (\text{rank}_G(\mathcal{Q}_{l'}) - \text{rank}_G(\mathcal{Q}_{l'-1})) - \text{rank}_G(\mathcal{Q}_0) \\
&\stackrel{(5)}{\geq} \left| \bigsqcup_{j=1}^{\rho-1} \mathcal{N}_j \right| - \text{rank}_G\left(\bigsqcup_{j=1}^{\rho-1} \mathcal{N}_j\right) + l(\delta - 1) \\
&= \sum_{j=1}^{\rho-1} (n_j - \xi_j) + \left( \left\lceil \frac{k - \sum_{j=1}^{\rho-1} \xi_j}{\rho} \right\rceil - 1 \right) (\delta - 1).
\end{aligned}$$

■

The next lemma gives a tight bound for the auxiliary parameter  $\xi$  with the original locality profile parameters  $(\mathbf{n}, \delta)$  appearing in the expression, again using Algorithm 1.

**Lemma 6.** *For linear  $[n, k]$  codes with locality profile  $(\mathbf{n}, \delta)$ , we have*

$$\xi_j \leq k_j,$$

for  $j \in [r^*]$ .

*Proof:* Considering the incremental symbols in the construction of  $Q_L \subset [n]$  in Algorithm 1, we obtain

$$\begin{aligned}
n_j &\geq |\mathcal{Q}_L| - |\mathcal{Q}_0| = \sum_{l=1}^L (|\mathcal{Q}_l| - |\mathcal{Q}_{l-1}|) \\
&\stackrel{(5)}{\geq} \sum_{l=1}^L (\text{rank}_G(\mathcal{Q}_l) - \text{rank}_G(\mathcal{Q}_{l-1})) + L(\delta - 1) \\
&\stackrel{(4)}{\geq} \xi_j + \left\lceil \frac{\xi_j}{j} \right\rceil (\delta - 1).
\end{aligned} \tag{6}$$

For  $0 \leq q_j \leq \delta - 2$ , suppose that  $\xi_j \geq p_j j + 1$ . It follows from (6) that

$$\begin{aligned}
n_j &\geq p_j j + 1 + (p_j + 1)(\delta - 1) \\
&= p_j(j + \delta - 1) + \delta \\
&> p_j(j + \delta - 1) + q_j \\
&= n_j,
\end{aligned}$$

which is a contradiction. Therefore, we have

$$\xi_j \leq p_j j = \lfloor m_j \rfloor j.$$

On the other hand, for  $\delta - 1 \leq q_j \leq j + \delta - 2$ , suppose that  $\xi_j \geq p_j j + q_j - (\delta - 1) + 1$ , hence  $\xi_j \geq p_j j + 1$ .

Again by (6), we have

$$\begin{aligned} n_j &\geq p_j j + q_j - (\delta - 1) + 1 + (p_j + 1)(\delta - 1) \\ &= p_j(j + \delta - 1) + q_j + 1 \\ &> n_j, \end{aligned}$$

and therefore

$$\begin{aligned} \xi_j &\leq p_j j + q_j - (\delta - 1) \\ &= n_j - \lceil m_j \rceil (\delta - 1). \end{aligned}$$

■

As a simple corollary to Lemma 6, we derive the following upper bound on the dimension of  $(r, \delta)$ -LRCs based on their locality profile parameters.

**Theorem 1** (Dimension upper bound for codes with unequal locality profile). *The dimension of linear  $[n, k]$  codes with locality profile  $(\mathbf{n}, \delta)$  is upper bounded by*

$$k \leq k_{UB}^{prf}(\mathbf{n}, \delta) \triangleq \sum_{j=1}^{r^*} k_j.$$

*Proof:* Clearly by Lemma 6, we have  $k = \sum_{j=1}^{r^*} \xi_j \leq \sum_{j=1}^{r^*} k_j$ . ■

Our minimum distance upper bound for  $(r, \delta)$ -LRCs with unequal locality profile, which is the main result of this section, is given in the following theorem.

**Theorem 2** (Minimum distance upper bound for codes with unequal locality profile). *The minimum Hamming distance of linear  $[n, k, d]$  codes with locality profile  $(\mathbf{n}, \delta)$  is upper bounded by*

$$d \leq d_{UB}^{prf}(\mathbf{n}, \delta) \triangleq n - k + 1 - \sum_{j=1}^{r-1} (n_j - k_j) - \left( \left\lceil \frac{k - \sum_{j=1}^{r-1} k_j}{r} \right\rceil - 1 \right) (\delta - 1),$$

where

$$r = \min \{j \in [r^*] \mid \sum_{j'=1}^j k_{j'} \geq k\}.$$

*Proof:* First note that  $r$  is well defined due to Theorem 1, and we have  $r \leq \rho$  since  $\sum_{j=1}^{\rho} k_i \geq \sum_{j=1}^{\rho} \xi_i = k$ . If  $r = \rho$ , it is easy to verify that the theorem holds by applying Lemma 6 on Lemma 5.

Otherwise, if  $r \leq \rho - 1$ , we get

$$\begin{aligned} d &\stackrel{(a)}{\leq} n - k + 1 - \sum_{j=1}^{\rho-1} (n_j - \xi_j) - \left( \left\lceil \frac{k - \sum_{j=1}^{\rho-1} \xi_j}{\rho} \right\rceil - 1 \right) (\delta - 1) \\ &\stackrel{(b)}{\leq} n - k + 1 - \sum_{j=1}^r (n_j - \xi_j) \\ &\stackrel{(c)}{\leq} n - k + 1 - \sum_{j=1}^{r-1} (n_j - k_j) - (n_r - k_r), \end{aligned} \tag{7}$$

where (a) is just Lemma 5, (b) is obtained by removing some non-negative subtrahends, and (c) is due to Lemma 6.

Note that, if  $0 \leq q_r \leq \delta - 2$ , we can write

$$\begin{aligned} n_r - k_r &= n_r - \lfloor m_r \rfloor r \geq n_r - m_r r \\ &= m_r(\delta - 1) \geq \lfloor m_r \rfloor(\delta - 1) \\ &= \frac{k_r}{r}(\delta - 1). \end{aligned} \tag{8}$$

Otherwise, if  $\delta - 1 \leq q_r \leq r + \delta - 2$ , again we get

$$\begin{aligned} n_r - k_r &= \lceil m_r \rceil(\delta - 1) \geq \frac{m_r r}{r}(\delta - 1) \\ &= \frac{n_r - m_r(\delta - 1)}{r}(\delta - 1) \geq \frac{n_r - \lceil m_r \rceil(\delta - 1)}{r}(\delta - 1) \\ &= \frac{k_r}{r}(\delta - 1). \end{aligned} \tag{9}$$

Furthermore, we have

$$\frac{k_r}{r} \geq \frac{k - \sum_{j=1}^{r-1} k_j}{r} > \left\lceil \frac{k - \sum_{j=1}^{r-1} k_j}{r} \right\rceil - 1. \tag{10}$$

Therefore, substituting (8), (9) and (10) into (7) completes the proof.  $\blacksquare$

For the conventional  $r$ -locality case, i.e.,  $\delta = 2$ , note that  $k_j = n_j - \lceil m_j \rceil(\delta - 1)$  regardless of  $q_j$ . Further substituting  $\delta = 2$  into Theorem 2 results in (2).

## V. UPPER BOUNDS BASED ON LOCALITY REQUIREMENT

The bounds derived in the previous section require knowing the locality profile. In a practical scenario, a fixed known locality profile setting is quite unlikely since there is no reason to exclude the locality profiles with better locality properties, i.e., smaller  $\text{loc}_\delta(\cdot)$  for some symbols, as long as the locality profile complies with the given locality requirement. Therefore, returning to the original problem formulation, we give two types of upper bounds based on the *locality requirement*.

First we establish upper bounds on the dimension and minimum distance of LRCs by making an exhaustive search through all locality profiles satisfying the given locality requirement. The lemma below characterizes the locality profile of a linear code that satisfy a given locality requirement. The resulting dimension and distance upper bounds are then given in a theorem following the lemma.

Note that, hereafter, we use a *hat* notation for locality profile-related parameters, such as  $\hat{\mathbf{n}}, \hat{n}_j, \hat{p}_j, \hat{q}_j, \hat{m}_j, \hat{k}_j$ , and  $\hat{r}$ , to distinguish them from their locality requirement counterparts.

**Lemma 7.** *A code satisfies the locality requirement  $(\mathbf{n}, \delta)$  if and only if its locality profile  $(\hat{\mathbf{n}}, \delta)$  is such that*

$$\sum_{j'=1}^j \hat{n}_{j'} \geq \sum_{j'=1}^j n_{j'},$$

for all  $j \in [r^*]$ .

*Proof:* See Appendix B. ■

**Theorem 3** (Upper bounds for codes with unequal locality requirement based on the locality profile). *The dimension and minimum Hamming distance of linear  $[n, k, d]$  codes satisfying the locality requirement  $(\mathbf{n}, \delta)$  are upper bounded by*

$$k \leq k_{UB}^{req(prf)}(\mathbf{n}, \delta) \triangleq \max_{\hat{\mathbf{n}} \in \mathcal{P}} \left\{ k_{UB}^{prf}(\hat{\mathbf{n}}, \delta) \right\},$$

$$d \leq d_{UB}^{req(prf)}(\mathbf{n}, \delta) \triangleq \max_{\hat{\mathbf{n}} \in \mathcal{P}} \left\{ d_{UB}^{prf}(\hat{\mathbf{n}}, \delta) \right\},$$

where

$$\mathcal{P} = \{ \hat{\mathbf{n}} \in \mathbb{Z}^{r^*} \mid \hat{n}_j \geq 0 \text{ and } \sum_{j'=1}^j \hat{n}_{j'} \geq \sum_{j'=1}^j n_{j'}, j \in [r^*] \}.$$

*Proof:* By Lemma 7, Definition 2, and Theorem 1 and 2, the theorem is self-evident. ■

Instead of the upper bounds in Theorem 3, it is desirable to have upper bounds in closed form, which will be given in Theorem 4 and 5. Although these bounds are, by construction, looser than Theorem 3 in general, it turns out that they coincide for some important parameter regimes.

Below, we first give a simple lemma followed by the two theorems stating the upper bounds. Moreover, the technique shown in its proof will also be utilized several times in the proof of the theorem stating the minimum distance upper bound.

**Lemma 8.** *For the locality profile  $(\hat{\mathbf{n}}, \delta)$  of a code satisfying the locality requirement  $(\mathbf{n}, \delta)$ , we have*

$$\sum_{j'=1}^j \hat{m}_{j'} \geq \sum_{j'=1}^j m_{j'}, \quad (11)$$

for all  $j \in [r^*]$ .

*Proof:* By repeatedly using Lemma 7, we have

$$\begin{aligned} \sum_{j'=1}^j \hat{m}_{j'} - \sum_{j'=1}^j m_{j'} &= \frac{\hat{n}_1 - n_1}{1 + \delta - 1} + \frac{\hat{n}_2 - n_2}{2 + \delta - 1} + \cdots + \frac{\hat{n}_j - n_j}{j + \delta - 1} \\ &\geq \frac{\sum_{j'=1}^2 (\hat{n}_{j'} - n_{j'})}{2 + \delta - 1} + \cdots + \frac{\hat{n}_j - n_j}{j + \delta - 1} \\ &\geq \cdots \geq \frac{\sum_{j'=1}^j (\hat{n}_{j'} - n_{j'})}{j + \delta - 1} \\ &\geq 0. \end{aligned}$$
■

**Theorem 4** (Dimension upper bound for codes with unequal locality requirement). *The dimension of linear  $[n, k, d]$  codes satisfying the locality requirement  $(\mathbf{n}, \delta)$  is upper bounded by*

$$k \leq k_{UB}^{req}(\mathbf{n}, \delta) \triangleq \sum_{j=1}^{r^*} m_j j.$$

**Algorithm 2** Used in the Proof of Theorem 5

---

```

1: Let  $j_0 = 0, l = 0, l' = 1$ 
2: while  $l' \leq \hat{r} - 1$  do
3:   if  $\sum_{j=j_l+1}^{l'} (\hat{p}_j - p_j + \hat{\phi}_j) < 0$  then
4:      $l = l + 1$ 
5:      $j_l = l'$ 
6:   end if
7:    $l' = l' + 1$ 
8: end while
9:  $L = l$ 

```

---

*Proof:* Let the locality profile of the code be  $(\hat{\mathbf{n}}, \delta)$ . We have

$$\begin{aligned}
k &\stackrel{(a)}{\leq} \sum_{j=1}^{r^*} \hat{k}_j \\
&\stackrel{(b)}{\leq} \sum_{j=1}^{r^*} \{\hat{n}_j - \hat{m}_j(\delta - 1)\} = n - (\delta - 1) \sum_{j=1}^{r^*} \hat{m}_j \\
&\stackrel{(c)}{\leq} \sum_{j=1}^{r^*} n_j - \sum_{j=1}^{r^*} m_j(\delta - 1) \\
&= \sum_{j=1}^{r^*} m_j j,
\end{aligned}$$

where (a) is Theorem 1, (b) is from Definition 2, and (c) is due to Lemma 8. ■

Note that Theorem 4 characterizes the feasible rate region for LRCs with an arbitrary locality requirement and degenerates to the rate bound in [24] (see also [14]) if all symbol locality requirements are equal.

**Theorem 5** (Minimum distance upper bound for codes with unequal locality requirement). *The minimum Hamming distance of linear  $[n, k, d]$  codes satisfying the locality requirement  $(\mathbf{n}, \delta)$  is upper bounded by*

$$d \leq d_{UB}^{req}(\mathbf{n}, \delta) \triangleq n - k + 1 - \sum_{j=1}^{r-1} \lfloor m_j \rfloor (\delta - 1) - \left( \left\lceil \frac{k - \sum_{j=1}^{r-1} \lfloor m_j \rfloor j}{r} \right\rceil - 1 \right) (\delta - 1),$$

where

$$r = \max \{0 \leq j \leq r^* - 1 \mid \sum_{j'=1}^j \lfloor m_{j'} \rfloor j' < k\} + 1.$$

*Proof:* Let the locality profile of the code be  $(\hat{\mathbf{n}}, \delta)$ . By Theorem 2, we have

$$d \leq n - k + 1 - \sum_{j=1}^{\hat{r}-1} (\hat{n}_j - \hat{k}_j) - \left( \left\lceil \frac{k - \sum_{j=1}^{\hat{r}-1} \hat{k}_j}{\hat{r}} \right\rceil - 1 \right) (\delta - 1). \quad (12)$$

For  $j \in [r^*]$ , let

$$\hat{\phi}_j = \begin{cases} 0 & \text{if } 0 \leq \hat{q}_j \leq \delta - 2, \\ 1 & \text{if } \delta - 1 \leq \hat{q}_j \leq j + \delta - 2. \end{cases}$$

Note that

$$\hat{k}_j = \hat{p}_j j + (\hat{q}_j - \delta + 1)\hat{\phi}_j, \quad (13)$$

and also

$$\hat{n}_j - \hat{k}_j = (\hat{p}_j + \hat{\phi}_j)(\delta - 1) + \hat{q}_j(1 - \hat{\phi}_j). \quad (14)$$

The proof will proceed with the corresponding cases.

*Case 1:  $r \geq \hat{r}$ .*

Substituting (13) and (14) into (12) yields

$$\begin{aligned} d &\leq n - k + 1 - \sum_{j=1}^{\hat{r}-1} \{(\hat{p}_j + \hat{\phi}_j)(\delta - 1) + \hat{q}_j(1 - \hat{\phi}_j)\} \\ &\quad - \left( \left\lceil \frac{k - \sum_{j=1}^{\hat{r}-1} \{\hat{p}_j j + (\hat{q}_j - \delta + 1)\hat{\phi}_j\}}{\hat{r}} \right\rceil - 1 \right) (\delta - 1) \\ &\leq n - k + 1 - \sum_{j=1}^{\hat{r}-1} p_j(\delta - 1) - \left( \left\lceil \frac{k - \sum_{j=1}^{\hat{r}-1} p_j j}{\hat{r}} \right\rceil - 1 + \lfloor A \rfloor + B \right) (\delta - 1), \end{aligned} \quad (15)$$

with

$$A = \frac{\sum_{j=1}^{\hat{r}-1} (\hat{p}_j - p_j)(\hat{r} - j) + \sum_{j=1}^{\hat{r}-1} (\hat{r} + \delta - 1 - \hat{q}_j)\hat{\phi}_j}{\hat{r}} \quad (16)$$

and

$$B = \frac{\sum_{j=1}^{\hat{r}-1} \hat{q}_j(1 - \hat{\phi}_j)}{\delta - 1}, \quad (17)$$

where (15) follows from the fact that  $\lceil a - b \rceil \geq \lceil a \rceil + \lfloor -b \rfloor$ .

Next, we will show that  $\lfloor A \rfloor + B \geq 0$ . Define  $j_l$ ,  $0 \leq l \leq L \leq \hat{r} - 1$ , according to Algorithm 2. Note that,  $j_0 = 0 < j_1 < \dots < j_L \leq \hat{r} - 1$  such that, for  $l \in [L]$ ,

$$\sum_{j=j_{l-1}+1}^{j'} (\hat{p}_j - p_j + \hat{\phi}_j) \geq 0, \quad (18)$$

$j_{l-1} + 1 \leq j' \leq j_l - 1$ , and

$$\sum_{j=j_{l-1}+1}^{j_l} (\hat{p}_j - p_j + \hat{\phi}_j) < 0. \quad (19)$$

Also, we have

$$\sum_{j=j_L+1}^{j'} (\hat{p}_j - p_j + \hat{\phi}_j) \geq 0, \quad (20)$$



$j_L + 1 \leq j' \leq r' - 1$ . Starting from (16), we can write

$$\begin{aligned}
A &\stackrel{(a)}{\geq} \frac{\sum_{j=1}^{\hat{r}-1} (\hat{p}_j - p_j + \hat{\phi}_j)(\hat{r} - j)}{\hat{r}} \\
&= \frac{\sum_{l=1}^L \sum_{j=j_{l-1}+1}^{j_l} (\hat{p}_j - p_j + \hat{\phi}_j)(\hat{r} - j) + \sum_{j=j_L+1}^{\hat{r}-1} (\hat{p}_j - p_j + \hat{\phi}_j)(\hat{r} - j)}{\hat{r}} \\
&\stackrel{(b)}{\geq} \frac{\sum_{l=1}^L \sum_{j=j_{l-1}+1}^{j_l} (\hat{p}_j - p_j + \hat{\phi}_j)(\hat{r} - j_l) + \sum_{j=j_L+1}^{\hat{r}-1} (\hat{p}_j - p_j + \hat{\phi}_j)}{\hat{r}} \\
&\geq \frac{\sum_{l=1}^L (\hat{r} - j_l) \sum_{j=j_{l-1}+1}^{j_l} (\hat{p}_j - p_j + \hat{\phi}_j)}{\hat{r}} \\
&\stackrel{(c)}{\geq} \sum_{l=1}^L \sum_{j=j_{l-1}+1}^{j_l} (\hat{p}_j - p_j + \hat{\phi}_j) \\
&= \sum_{j=1}^{j_L} (\hat{p}_j - p_j + \hat{\phi}_j),
\end{aligned} \tag{21}$$

where (a) is due to the fact that  $\hat{q}_j < j + \delta - 1$ , (b) follows from (18) and (20) with a derivation similar to Lemma 8, and (c) comes from (19) also with a derivation similar to Lemma 8. As an intermediate result, we get

$$[A] \geq \sum_{j=1}^{j_L} (\hat{p}_j - p_j + \hat{\phi}_j).$$

For term B, we claim that

$$\begin{aligned}
B &\stackrel{(a)}{\geq} \frac{\sum_{j=1}^{j_L} \hat{q}_j (1 - \hat{\phi}_j)}{\delta - 1} \\
&\stackrel{(b)}{\geq} - \sum_{j=1}^{j_L} (\hat{p}_j - p_j + \hat{\phi}_j),
\end{aligned}$$

where (a) is obtained by removing some non-negative summands from (17). To see why (b) is true, first note that Lemma 7 implies

$$\sum_{j=1}^{j_L} \hat{q}_j \geq \sum_{j=1}^{j_L} q_j - \sum_{j=1}^{j_L} (\hat{p}_j - p_j)(j + \delta - 1).$$

Therefore, we have

$$\begin{aligned}
\sum_{j=1}^{j_L} \hat{q}_j (1 - \hat{\phi}_j) &\geq \sum_{j=1}^{j_L} q_j - \sum_{j=1}^{j_L} (\hat{p}_j - p_j)(j + \delta - 1) - \sum_{j=1}^{j_L} \hat{q}_j \hat{\phi}_j \\
&\stackrel{(a)}{\geq} - \sum_{j=1}^{j_L} (\hat{p}_j - p_j + \hat{\phi}_j)(j + \delta - 1) \\
&= - \sum_{l=1}^L \sum_{j=j_{l-1}+1}^{j_l} (\hat{p}_j - p_j + \hat{\phi}_j)(j + \delta - 1) \\
&\stackrel{(b)}{\geq} - \sum_{l=1}^L (j_l + \delta - 1) \sum_{j=j_{l-1}+1}^{j_l} (\hat{p}_j - p_j + \hat{\phi}_j) \\
&\stackrel{(c)}{\geq} -(\delta - 1) \sum_{j=1}^{j_L} (\hat{p}_j - p_j + \hat{\phi}_j),
\end{aligned}$$

where (a), (b), and (c) follows from observations identical to those in (21), respectively.

Since  $\lfloor A \rfloor + B \geq 0$ , we can write, continuing from (15),

$$\begin{aligned}
d &\leq n - k + 1 - \sum_{j=1}^{\hat{r}-1} p_j(\delta - 1) - \left( \left\lceil \frac{k - \sum_{j=1}^{\hat{r}-1} p_j j}{\hat{r}} \right\rceil - 1 \right) (\delta - 1) \\
&= n - k + 1 - \sum_{j=1}^{r-1} p_j(\delta - 1) - \left( \left\lceil \frac{k - \sum_{j=1}^{\hat{r}-1} p_j j - \sum_{j=\hat{r}}^{r-1} p_j \hat{r}}{\hat{r}} \right\rceil - 1 \right) (\delta - 1) \\
&\leq n - k + 1 - \sum_{j=1}^{r-1} p_j(\delta - 1) - \left( \left\lceil \frac{k - \sum_{j=1}^{r-1} p_j j}{r} \right\rceil - 1 \right) (\delta - 1) \\
&= n - k + 1 - \sum_{j=1}^{r-1} \lfloor m_j \rfloor (\delta - 1) - \left( \left\lceil \frac{k - \sum_{j=1}^{r-1} \lfloor m_j \rfloor j}{r} \right\rceil - 1 \right) (\delta - 1).
\end{aligned}$$

Case 2:  $r < \hat{r}$ .

It is easy to verify that (14) implies  $\hat{n}_j - \hat{k}_j \geq \hat{m}_j(\delta - 1)$ ,  $j \in [r^*]$ . By applying Lemma 8, we can write

$$\sum_{j=1}^r (\hat{n}_j - \hat{k}_j) \geq \sum_{j=1}^r \hat{m}_j(\delta - 1) \geq \sum_{j=1}^r \lfloor m_j \rfloor (\delta - 1). \quad (22)$$

On the other hand, since  $r \leq \hat{r} - 1 \leq r^* - 1$ , we have  $\sum_{j=1}^r \lfloor m_j \rfloor j \geq k$ , leading to

$$\lfloor m_r \rfloor \geq \frac{k - \sum_{j=1}^{r-1} \lfloor m_j \rfloor j}{r} > \left\lceil \frac{k - \sum_{j=1}^{r-1} \lfloor m_j \rfloor j}{r} \right\rceil - 1. \quad (23)$$

Therefore, from (12), we get

$$\begin{aligned}
d &\stackrel{(a)}{\leq} n - k + 1 - \sum_{j=1}^r (\hat{n}_j - \hat{k}_j) \\
&\stackrel{(b)}{\leq} n - k + 1 - \sum_{j=1}^{r-1} \lfloor m_j \rfloor (\delta - 1) - \left( \left\lceil \frac{k - \sum_{j=1}^{r-1} \lfloor m_j \rfloor j}{r} \right\rceil - 1 \right) (\delta - 1),
\end{aligned}$$

where (a) is obtained by removing some non-negative subtrahends, and (b) follows from (22) and (23).  $\blacksquare$

**Remark 4.** Note that for  $r \leq r' \leq r^*$ ,

$$\begin{aligned}
\sum_{j=1}^{r-1} \lfloor m_j \rfloor + \left\lceil \frac{k - \sum_{j=1}^{r-1} \lfloor m_j \rfloor j}{r} \right\rceil &= \sum_{j=1}^{r'-1} \lfloor m_j \rfloor + \left\lceil \frac{k - \sum_{j=1}^{r-1} \lfloor m_j \rfloor j - \sum_{j=r}^{r'-1} \lfloor m_j \rfloor r}{r} \right\rceil \\
&\geq \sum_{j=1}^{r'-1} \lfloor m_j \rfloor + \left\lceil \frac{k - \sum_{j=1}^{r'-1} \lfloor m_j \rfloor j}{r'} \right\rceil.
\end{aligned}$$

Therefore, we also have

$$d \leq n - k + 1 - \sum_{j=1}^{r^*-1} \lfloor m_j \rfloor (\delta - 1) - \left( \left\lceil \frac{k - \sum_{j=1}^{r^*-1} \lfloor m_j \rfloor j}{r^*} \right\rceil - 1 \right) (\delta - 1),$$

which is somewhat looser than Theorem 5, but does not require the computation of  $r$ . A similar observation has been made implicitly in [9].

The next two corollaries show the relationship between the upper bounds based on the evaluation of the relevant locality profiles (Theorem 3) and the upper bounds in closed form (Theorem 4 and 5).

**Corollary 1.** *For the upper bounds on codes satisfying a given locality requirement  $(\mathbf{n}, \delta)$  (Theorem 3, 4, and 5), we have*

$$\begin{aligned} k_{UB}^{req(prf)}(\mathbf{n}, \delta) &\leq k_{UB}^{req}(\mathbf{n}, \delta), \\ d_{UB}^{req(prf)}(\mathbf{n}, \delta) &\leq d_{UB}^{req}(\mathbf{n}, \delta). \end{aligned}$$

*Proof:* From the proofs of Theorem 4 and 5, we have

$$\begin{aligned} k_{UB}^{prf}(\hat{\mathbf{n}}, \delta) &\leq k_{UB}^{req}(\mathbf{n}, \delta), \\ d_{UB}^{prf}(\hat{\mathbf{n}}, \delta) &\leq d_{UB}^{req}(\mathbf{n}, \delta), \end{aligned}$$

for an arbitrary locality profile  $(\hat{\mathbf{n}}, \delta)$  that satisfies the locality requirement  $(\mathbf{n}, \delta)$ . In other words,  $(\hat{\mathbf{n}}, \delta) \in \mathcal{P}$ , where  $\mathcal{P}$  is defined in Theorem 3. Therefore, the corollary simply follows from Theorem 3.  $\blacksquare$

**Corollary 2.** *For the locality requirement  $(\mathbf{n}, \delta)$  such that  $j + \delta - 1 \mid n_j$ ,  $j \in [r^*]$ , we have*

$$\begin{aligned} k_{UB}^{req(prf)}(\mathbf{n}, \delta) &= k_{UB}^{req}(\mathbf{n}, \delta), \\ d_{UB}^{req(prf)}(\mathbf{n}, \delta) &= d_{UB}^{req}(\mathbf{n}, \delta). \end{aligned}$$

*Proof:* Note that

$$\begin{aligned} k_{UB}^{prf}(\mathbf{n}, \delta) &\stackrel{(a)}{\leq} k_{UB}^{req(prf)}(\mathbf{n}, \delta) \stackrel{(b)}{\leq} k_{UB}^{req}(\mathbf{n}, \delta), \\ d_{UB}^{prf}(\mathbf{n}, \delta) &\stackrel{(a)}{\leq} d_{UB}^{req(prf)}(\mathbf{n}, \delta) \stackrel{(b)}{\leq} d_{UB}^{req}(\mathbf{n}, \delta), \end{aligned}$$

where (a) is due to the definitions of  $k_{UB}^{req(prf)}$  and  $d_{UB}^{req(prf)}$ , and (b) is just Corollary 1. The proof is complete by checking that

$$\begin{aligned} k_{UB}^{prf}(\mathbf{n}, \delta) &= k_{UB}^{req}(\mathbf{n}, \delta), \\ d_{UB}^{prf}(\mathbf{n}, \delta) &= d_{UB}^{req}(\mathbf{n}, \delta), \end{aligned}$$

with the condition of  $j + \delta - 1 \mid n_j$ , which results in  $m_j = \lceil m_j \rceil = \lfloor m_j \rfloor$  and  $k_j = n_j - m_j(\delta - 1) = m_j j$ . The equality between  $d_{UB}^{prf}(\mathbf{n}, \delta)$  and  $d_{UB}^{req}(\mathbf{n}, \delta)$  can be shown as

$$\begin{aligned} r^{req}(\mathbf{n}, \delta) &= \max \{0 \leq j \leq r^* - 1 \mid \sum_{j'=1}^j m_{j'} j' < k\} + 1 \\ &\stackrel{(a)}{=} \min \{j \in [r^*] \mid \sum_{j'=1}^j m_{j'} j' \geq k\} \\ &= r^{prf}(\mathbf{n}, \delta), \end{aligned}$$

where (a) is due to Theorem 4.  $\blacksquare$

## VI. OPTIMAL CONSTRUCTION

We give an optimal code construction achieving the equality for the bound in Theorem 5, and also the bound in Theorem 3. In other words, the code is built in the parameter regime where the two distance bounds coincide, as shown in Corollary 2. The construction closely follows the Gabidulin-based LRC construction which originates from [13], and is also used in [7].

**Construction 1** (Gabidulin-based LRC with unequal locality). *For integers  $m_j$ ,  $j \in [r^*]$ , let  $n_j = m_j(j + \delta - 1)$  and  $n = \sum_{j=1}^{r^*} n_j$ . Let us also constrain the parameters to satisfy the condition  $r^* \leq k \leq n_{\text{Gab}} \triangleq \sum_{j=1}^{r^*} m_j j \leq t$ . Linear  $[n, k]_{q^t}$  codes satisfying the locality requirement  $(\mathbf{n}, \delta)$  are constructed according to the following steps.*

- 1) *Precode  $k$  information symbols using a Gabidulin code of length  $n_{\text{Gab}}$ .*
- 2) *Partition the Gabidulin codeword symbols into  $\sum_{j=1}^{r^*} m_j$  local groups, where each  $m_j$  groups are of size  $j$ ,  $j \in [r^*]$ .*
- 3) *Encode each local group of size  $j$  using an  $[j + \delta - 1, j, \delta]_q$  MDS code.*

Note that according to Lemma 4, every symbol of the code by Construction 1 also corresponds to an evaluation of the polynomial used in the Gabidulin precoding.

**Remark 5.** *Clearly, the subspace generated by the evaluation points of the code of Construction 1 is a direct sum of each subspace generated by the evaluation points corresponding to a single local group. Therefore,  $\text{rank}_{\mathbb{E}}(\mathcal{T})$  of some set  $\mathcal{T} \subset [n]$  is the sum of each  $\text{rank}_{\mathbb{E}}(\cdot)$  computed separately on the points being in the same local group.*

To analyze the minimum distance of the code by Construction 1, we require a lemma that can be understood quite intuitively. The optimality of the code construction is claimed in the theorem following the lemma.

**Lemma 9.** *Suppose an ordered set  $\mathcal{L} = \{\mathcal{G}_1, \dots, \mathcal{G}_{|\mathcal{L}|}\}$  such that  $|\mathcal{L}| = \sum_{j=1}^{r^*} m_j$ , where each element of  $\mathcal{L}$  is a symbol index set corresponding to the symbols of a distinct encoded local group in Construction 1, and the order is according to the size of the local groups in ascending order. Local groups of identical size are ordered arbitrarily. For codes by Construction 1, the worst case erasure pattern for  $e$  erased symbols, in terms of rank erasure, corresponds to the case where the  $n - e$  remaining symbols are taken greedily, starting from the first element  $\mathcal{G}_1$  of  $\mathcal{L}$ .*

*Proof:* See Appendix C. ■

**Theorem 6** (Optimality of Gabidulin-based LRC construction). *Gabidulin-based LRCs (Construction 1) satisfy the locality requirement  $(\mathbf{n}, \delta)$ , and the equality is achieved in the distance bound for codes with unequal locality requirement (Theorem 5).*

*Proof:* It is obvious by construction that a Gabidulin-based LRC  $\mathcal{C}$  satisfies the locality requirement  $(\mathbf{n}, \delta)$ . In particular, by choosing  $\mathcal{S}_i$  as the index set of symbols that belongs to the same MDS local code, we have  $i \in \mathcal{S}_i$  and  $|\mathcal{S}_i| = j + \delta - 1$ . Furthermore,  $d(\mathcal{C}|_{\mathcal{S}_i}) \geq \delta$  since  $\mathcal{C}|_{\mathcal{S}_i}$  is a subcode of a  $[j + \delta - 1, j, \delta]_q$  MDS code.

To analyze the minimum distance, we give a lower bound on the minimum distance of the code, which equals the upper bound of Theorem 5. In particular, we derive a lower bound by using Lemma 3 as well as Remark 1. Note that erasure correction is possible from an arbitrary set with the cardinality of

$$\tau = k + \sum_{j=1}^{r-1} m_j(\delta - 1) + \left( \left\lceil \frac{k - \sum_{j=1}^{r-1} m_j j}{r} \right\rceil - 1 \right) (\delta - 1),$$

where  $r$  is given by Theorem 5.

Let integers  $P$  and  $Q$  such that

$$k - 1 - \sum_{j=1}^{r-1} m_j j = Pr + Q \geq 0 \quad (24)$$

and  $0 \leq Q \leq r - 1$ . Consider an arbitrary symbol index set  $\mathcal{T} \subset [n]$  of cardinality

$$|\mathcal{T}| = \sum_{j=1}^{r-1} n_j + P(r + \delta - 1) + Q + 1. \quad (25)$$

According to Lemma 9, the number of *rank erasures* is maximal when  $\mathcal{T}$  corresponds to all symbols in every local group of size (before MDS encoding) at most  $r - 1$  and  $P$  local groups of size  $r$ , and some  $Q + 1$  symbols in an additional local group of size  $r$ . Such a distribution for  $\mathcal{T}$  is valid since

$$\begin{aligned} P &\stackrel{(24)}{=} \frac{k - \sum_{j=1}^{r-1} m_j j}{r} - \frac{1 + Q}{r} \\ &\stackrel{(a)}{\leq} \frac{m_r r}{r} - \frac{1 + Q}{r} \\ &< m_r, \end{aligned}$$

where (a) comes from the definition of  $r$  if  $r < r^*$ , or from the parameter condition in the construction if  $r = r^*$ .

By Lemma 4 and Remark 5, we have

$$\text{rank}_E(\mathcal{T}) \geq \sum_{j=1}^{r-1} m_j j + Pr + Q + 1 = k,$$

and erasure correction is therefore possible from arbitrary  $\mathcal{T}$ .

The proof is complete by noting that substituting (24) into (25) yields

$$|\mathcal{T}| = k + \sum_{j=1}^{r-1} m_j(\delta - 1) + P(\delta - 1),$$

where

$$P \stackrel{(24)}{=} \left\lceil \frac{k - \sum_{j=1}^{r-1} m_j j - 1}{r} \right\rceil = \left\lceil \frac{k - \sum_{j=1}^{r-1} m_j j}{r} \right\rceil - 1,$$

which is equal to  $\tau$ . ■

## VII. FURTHER RESULTS

### A. Optimality in terms of the Profile-based Bound

In this subsection, we show that Construction 1 is also optimal in terms of the distance upper bound based on the *locality profile* (Theorem 2). Clearly, as mentioned in the proof of Corollary 2, the minimum distance of codes

by Construction 1 that satisfy the *locality requirement*  $(\mathbf{n}, \delta)$  achieve the equality for the bound in Theorem 2, i.e.,  $d = d_{UB}^{req}(\mathbf{n}, \delta) = d_{UB}^{prf}(\mathbf{n}, \delta)$ . To claim optimality, we have to show that the locality profile of the constructed code is the same as the locality requirement  $(\mathbf{n}, \delta)$ , which is given as a proposition after the following lemma.

**Lemma 10.** *For a symbol index set  $\mathcal{T} \subset [n]$  of an  $[n, k]$  Gabidulin-based LRC (Construction 1), such that  $\text{rank}_G(\mathcal{T}) < k$ , we have*

$$\text{rank}_E(\mathcal{T}) = \text{rank}_G(\mathcal{T}).$$

*Proof:* See Appendix D. ■

**Proposition 1.** *A Gabidulin-based LRC (Construction 1) has a locality profile  $(\mathbf{n}, \delta)$ , which equals the locality requirement used in the construction, and therefore are optimal in terms of the distance bound for codes with unequal locality profile (Theorem 2).*

*Proof:* For  $j \in [r^*]$ , let  $\mathcal{N}_j$  denote the symbol index set of the Gabidulin-based LRC  $\mathcal{C}$  corresponding to the local groups of size (before MDS encoding)  $j$ . We have to show that  $\text{loc}_\delta(i) = j$  for every  $i \in \mathcal{N}_j$ . Since it is obvious by construction that a Gabidulin-based LRC  $\mathcal{C}$  satisfies the locality requirement  $\text{loc}_\delta(i) \leq j$  (see the proof of Theorem 6), we further show that assuming  $\text{loc}_\delta(i) < j$  for  $i \in \mathcal{N}_j$  results in a contradiction, thereby completing the proof.

Suppose that there exists some  $i \in \mathcal{N}_j$  with  $\text{loc}_\delta(i) = j'$  such that  $j' < j$ . This implies the existence of a set  $\mathcal{S}'_i$  such that  $i \in \mathcal{S}'_i$  and

$$|\mathcal{S}'_i| = j' + \delta - 1 < j + \delta - 1,$$

with  $d(\mathcal{C}|_{\mathcal{S}'_i}) = \delta$ . We claim that for an arbitrary set  $\mathcal{T} \subset \mathcal{S}'_i$  such that  $|\mathcal{T}| \geq |\mathcal{S}'_i| - (\delta - 1)$ , it must be true that

$$\text{rank}_E(\mathcal{T}) = \text{rank}_E(\mathcal{S}'_i). \quad (26)$$

First, note that  $\mathcal{T}$  is an erasure correctable set for  $\mathcal{C}|_{\mathcal{S}'_i}$ , hence  $\text{rank}_G(\mathcal{T}) = \dim(\mathcal{C}|_{\mathcal{S}'_i}) = \text{rank}_G(\mathcal{S}'_i)$  (see Remark 1). From the Singleton bound, we have

$$\text{rank}_G(\mathcal{S}'_i) \leq |\mathcal{S}'_i| - d(\mathcal{C}|_{\mathcal{S}'_i}) + 1 < j \leq r^* \leq k.$$

Therefore we can apply Lemma 10 and the claim (26) follows. The remaining part of the proof proceeds with two cases.

For the first case, assume that  $1 \leq |\mathcal{S}'_i \cap \mathcal{S}_i| \leq \delta - 1$  and let  $\mathcal{T} = \mathcal{S}'_i \setminus \mathcal{S}_i$ . Since we have  $i \in \mathcal{S}'_i$  and  $i \notin \mathcal{T}$  while  $\mathcal{T} \subset \mathcal{S}'_i$ , it follows that

$$\text{rank}_E(\mathcal{S}'_i) \geq \text{rank}_E(\mathcal{T} \sqcup \{i\}) \stackrel{(a)}{=} \text{rank}_E(\mathcal{T}) + 1,$$

which is a contradiction to (26). Note that (a) is due to Remark 5.

For the second case where  $\delta \leq |\mathcal{S}'_i \cap \mathcal{S}_i| < j + \delta - 1$ , let  $\mathcal{Q}$  and  $\mathcal{Q}'$  be arbitrary sets such that  $\mathcal{Q} \subset \mathcal{S}'_i \cap \mathcal{S}_i$  with  $|\mathcal{Q}| = \delta - 1$ , and  $\mathcal{Q}' \subsetneq \mathcal{Q}$ . By letting  $\mathcal{T} = \mathcal{S}'_i \setminus \mathcal{Q}$ , we get

$$\begin{aligned} \text{rank}_{\mathbb{E}}(\mathcal{T}) &= \text{rank}_{\mathbb{E}}((\mathcal{S}'_i \setminus \mathcal{S}_i) \sqcup ((\mathcal{S}'_i \cap \mathcal{S}_i) \setminus \mathcal{Q})) \\ &\stackrel{(a)}{=} \text{rank}_{\mathbb{E}}(\mathcal{S}'_i \setminus \mathcal{S}_i) + \text{rank}_{\mathbb{E}}((\mathcal{S}'_i \cap \mathcal{S}_i) \setminus \mathcal{Q}) \\ &\stackrel{(b)}{=} \text{rank}_{\mathbb{E}}(\mathcal{S}'_i \setminus \mathcal{S}_i) + |\mathcal{S}'_i \cap \mathcal{S}_i| - |\mathcal{Q}|, \end{aligned}$$

where (a) is due to Remark 5, and (b) follows from Lemma 4 with the observation that  $(\mathcal{S}'_i \cap \mathcal{S}_i) \setminus \mathcal{Q} \subset \mathcal{S}_i$  and  $|(\mathcal{S}'_i \cap \mathcal{S}_i) \setminus \mathcal{Q}| = |\mathcal{S}'_i \cap \mathcal{S}_i| - |\mathcal{Q}| \leq j$ . Similarly, setting  $\mathcal{T}' = \mathcal{S}'_i \setminus \mathcal{Q}'$ , we can write

$$\begin{aligned} \text{rank}_{\mathbb{E}}(\mathcal{T}') &= \text{rank}_{\mathbb{E}}((\mathcal{S}'_i \setminus \mathcal{S}_i) \sqcup ((\mathcal{S}'_i \cap \mathcal{S}_i) \setminus \mathcal{Q}')) \\ &= \text{rank}_{\mathbb{E}}(\mathcal{S}'_i \setminus \mathcal{S}_i) + \text{rank}_{\mathbb{E}}((\mathcal{S}'_i \cap \mathcal{S}_i) \setminus \mathcal{Q}') \\ &= \text{rank}_{\mathbb{E}}(\mathcal{S}'_i \setminus \mathcal{S}_i) + |\mathcal{S}'_i \cap \mathcal{S}_i| - |\mathcal{Q}'|. \end{aligned}$$

Therefore, we have

$$\text{rank}_{\mathbb{E}}(\mathcal{S}'_i) \geq \text{rank}_{\mathbb{E}}(\mathcal{T}') > \text{rank}_{\mathbb{E}}(\mathcal{T}),$$

which is again a contradiction to (26). ■

### B. Two Unequal Locality Requirement Case

We show that the closed form minimum distance upper bound based on the locality requirement (Theorem 5) can be tightened in the case where there are only two different locality requirements and a special condition holds. In particular, consider a locality requirement  $(\mathbf{n}, \delta)$  such that  $n_{j_1} \neq 0$  for some  $j_1 \in [r^* - 1]$  and  $n_j = 0$  for all  $j \in [r^* - 1] \setminus j_1$ . Note that by definition, we have  $n_{r^*} \neq 0$ . Letting  $j_2 = r^*$ , we use the notation  $((n_{j_1}, j_1), (n_{j_2}, j_2), \delta)$  for such a locality requirement. The tightened bound is given as a proposition after the following corollary restating Theorem 5 in a relevant form.

**Corollary 3.** *The minimum Hamming distance of linear  $[n, k, d]$  codes satisfying the locality requirement  $((n_{j_1}, j_1), (n_{j_2}, j_2), \delta)$  is upper bounded by*

1) if  $\lfloor m_{j_1} \rfloor j_1 \geq k$ ,

$$d \leq n - k + 1 - \left( \left\lceil \frac{k}{j_1} \right\rceil - 1 \right) (\delta - 1),$$

2) if  $\lfloor m_{j_1} \rfloor j_1 < k$ ,

$$d \leq n - k + 1 - \lfloor m_{j_1} \rfloor (\delta - 1) - \left( \left\lceil \frac{k - \lfloor m_{j_1} \rfloor j_1}{j_2} \right\rceil - 1 \right) (\delta - 1).$$

**Proposition 2.** *The minimum Hamming distance of linear  $[n, k, d]$  codes satisfying the locality requirement  $((n_{j_1}, j_1), (n_{j_2}, j_2), \delta)$ , such that  $q_{j_1} = 0$  or  $\delta - 1 \leq q_{j_1} \leq j_1 + \delta - 2$ , is upper bounded by*

1) if  $\lceil m_{j_1} \rceil j_1 \geq k$ ,

$$d \leq n - k + 1 - \left( \left\lceil \frac{k}{j_1} \right\rceil - 1 \right) (\delta - 1), \tag{27}$$

2) if  $\lceil m_{j_1} \rceil j_1 < k$ ,

$$d \leq n - k + 1 - \lceil m_{j_1} \rceil (\delta - 1) - \left( \left\lceil \frac{k - \lceil m_{j_1} \rceil j_1}{j_2} \right\rceil - 1 \right) (\delta - 1). \quad (28)$$

*Proof:* To derive a bound tighter than Corollary 3, we first make some auxiliary definitions. Specifically, for

$$\Delta \mathcal{N} = \{i \in \mathcal{N}_{j_2} \mid \text{loc}_\delta(i) \leq j_1\},$$

let

$$\tilde{\mathcal{N}}_{j_1} = \mathcal{N}_{j_1} \sqcup \Delta \mathcal{N},$$

$$\tilde{\mathcal{N}}_{j_2} = \mathcal{N}_{j_2} \setminus \Delta \mathcal{N},$$

and  $\tilde{n}_{j_i} = |\tilde{\mathcal{N}}_{j_i}|$ ,  $i = 1, 2$ . Furthermore,  $\check{p}_{j_1}$ ,  $\check{q}_{j_1}$  and  $\check{m}_{j_1}$  are defined such as in Definition 2 and 3. By definition, we have  $\check{n}_{j_1} \geq n_{j_1}$  and also

$$\check{m}_{j_1} \geq m_{j_1}. \quad (29)$$

Note that the sets  $\tilde{\mathcal{N}}_{j_1}$  and  $\tilde{\mathcal{N}}_{j_2}$  share an important property with those sets corresponding to the locality profile, such that their actual symbol localities are disjoint, i.e.,  $\text{loc}_\delta(i_1) \leq j_1$  and  $j_1 \leq \text{loc}_\delta(i_2) \leq j_2$  for  $i_1 \in \tilde{\mathcal{N}}_{j_1}$  and  $i_2 \in \tilde{\mathcal{N}}_{j_2}$ . Therefore, by repeating the steps in the proof of Lemma 5 and 6, and Theorem 2 on  $\tilde{\mathcal{N}}_{j_1}$  and  $\tilde{\mathcal{N}}_{j_2}$ , one can verify that

$$d \leq n - k + 1 - \gamma_0, \quad (30)$$

where

$$\gamma_0 = \begin{cases} \left( \left\lceil \frac{k}{j_1} \right\rceil - 1 \right) (\delta - 1) & \text{if } \check{k}_{j_1} \geq k, \\ \check{n}_{j_1} - \check{k}_{j_1} + \left( \left\lceil \frac{k - \check{k}_{j_1}}{j_2} \right\rceil - 1 \right) (\delta - 1) & \text{if } \check{k}_{j_1} < k, \end{cases} \quad (31a)$$

$$(31b)$$

and

$$\check{k}_{j_1} = \begin{cases} \lfloor \check{m}_{j_1} \rfloor j_1 & \text{if } 0 \leq \check{q}_{j_1} \leq \delta - 2, \\ \check{n}_{j_1} - \lceil \check{m}_{j_1} \rceil (\delta - 1) & \text{if } \delta - 1 \leq \check{q}_{j_1} \leq j_1 + \delta - 2. \end{cases} \quad (32)$$

Since the bound (27), being identical with (31a), is tighter than (28) (see Remark 4), the claim is valid if  $\check{k}_{j_1} \geq k$ , and we only have to check for the case where  $\check{k}_{j_1} < k$ .

1)  $\lceil m_{j_1} \rceil j_1 \geq k$ .

If  $0 \leq \check{q}_{j_1} \leq \delta - 2$ , then

$$\begin{aligned} \check{n}_{j_1} - \check{k}_{j_1} &\stackrel{(32)}{=} \check{n}_{j_1} - \lfloor \check{m}_{j_1} \rfloor j_1 \\ &\geq \check{n}_{j_1} - \check{m}_{j_1} j_1 = \check{m}_{j_1} (\delta - 1) \\ &\stackrel{(a)}{\geq} \lceil m_{j_1} \rceil (\delta - 1), \end{aligned} \quad (33)$$

where (a) is due to (29) and the conditions imposed on  $\check{q}_{j_1}$  and  $q_{j_1}$ .



If otherwise  $\delta - 1 \leq \check{q}_{j_1} \leq j_1 + \delta - 2$ , again we have (33) as

$$\check{n}_{j_1} - \check{k}_{j_1} \stackrel{(32)}{=} \lceil \check{m}_{j_1} \rceil (\delta - 1) \stackrel{(29)}{\geq} \lceil m_{j_1} \rceil (\delta - 1).$$

Furthermore, from the condition  $\lceil m_{j_1} \rceil j_1 \geq k$ , we have

$$\lceil m_{j_1} \rceil \geq \frac{k}{j_1} \geq \left\lceil \frac{k}{j_1} \right\rceil - 1. \quad (34)$$

Observing that  $\gamma_0 \geq \check{n}_{j_1} - \check{k}_{j_1}$  from (31b) and substituting into (30) with (33) and (34) results in (27).

2)  $\lceil m_{j_1} \rceil j_1 < k$ .

If  $0 \leq \check{q}_{j_1} \leq \delta - 2$ , we have

$$\begin{aligned} \check{n}_{j_1} - \check{k}_{j_1} &\stackrel{(32)}{=} \check{n}_{j_1} - \lfloor \check{m}_{j_1} \rfloor j_1 \geq \check{n}_{j_1} - \check{m}_{j_1} j_1 = \check{m}_{j_1} (\delta - 1) \\ &\geq \lfloor \check{m}_{j_1} \rfloor (\delta - 1). \end{aligned}$$

Substituting into (31b) together with (32), we get

$$\begin{aligned} \gamma_0 &\geq \lfloor \check{m}_{j_1} \rfloor (\delta - 1) + \left( \left\lceil \frac{k - \lfloor \check{m}_{j_1} \rfloor j_1}{j_2} \right\rceil - 1 \right) (\delta - 1) \\ &\stackrel{(a)}{\geq} \lceil m_{j_1} \rceil (\delta - 1) + \left( \left\lceil \frac{k - \lceil m_{j_1} \rceil j_1}{j_2} \right\rceil - 1 \right) (\delta - 1), \end{aligned}$$

where (a) holds since for integers  $a, b$  such that  $a \geq b$ , we have

$$\begin{aligned} a + \left\lceil \frac{k - ar_1}{r_2} \right\rceil &= b + \left\lceil \frac{k - br_1}{r_2} + (a - b) \left( 1 - \frac{r_1}{r_2} \right) \right\rceil \\ &\geq b + \left\lceil \frac{k - br_1}{r_2} \right\rceil, \end{aligned}$$

and it is true that  $\lfloor \check{m}_{j_1} \rfloor \geq \lceil m_{j_1} \rceil$  for the same reason as in (33). Now, (28) is immediate by (30).

Otherwise, if  $\delta - 1 \leq \check{q}_{j_1} \leq j_1 + \delta - 2$ , we have

$$\begin{aligned} \check{k}_{j_1} &\stackrel{(32)}{=} \check{n}_{j_1} - \lceil \check{m}_{j_1} \rceil (\delta - 1) \\ &\leq \check{n}_{j_1} - \check{m}_{j_1} (\delta - 1) = \check{m}_{j_1} j_1 \\ &\leq \lceil \check{m}_{j_1} \rceil j_1. \end{aligned}$$

Again, substituting into (31b) together with (32), we get

$$\begin{aligned} \gamma_0 &\geq \lceil \check{m}_{j_1} \rceil (\delta - 1) + \left( \left\lceil \frac{k - \lceil \check{m}_{j_1} \rceil j_1}{j_2} \right\rceil - 1 \right) (\delta - 1) \\ &\stackrel{(29)}{\geq} \lceil m_{j_1} \rceil (\delta - 1) + \left( \left\lceil \frac{k - \lceil m_{j_1} \rceil j_1}{j_2} \right\rceil - 1 \right) (\delta - 1), \end{aligned}$$

which concludes the proof after being substituted into (30). ■

It is easy to verify that Proposition 2 implies Corollary 3, hence a tighter bound. Note that in the parameter regime of our optimal code construction (Construction 1), Proposition 2 coincides with both Corollary 3 and the minimum distance of the optimal code construction.

For the special case of  $\delta = 2$ , the condition of Proposition 2, such that  $q_{j_1} = 0$  or  $\delta - 1 \leq q_{j_1} \leq j_1 + \delta - 2$ , is always true and therefore can be omitted. In this case, Proposition 2 becomes identical to [9, Thm. 6].<sup>3</sup> However, note that the proof of [9, Thm. 6] makes an implicit assumption that the given parameters  $n_{j_1}$  and  $n_{j_2}$  are not locality requirements, but in fact the profile-like parameters  $\check{n}_{j_1}$  and  $\check{n}_{j_2}$  used in the proof of Proposition 2. On the contrary, Proposition 2 is simply based on the locality requirement.

## VIII. CONCLUSION

In this work, we have investigated the minimum distance characteristics of  $(r, \delta)$ -LRCs with unequal locality. The problem has been analyzed in both the locality profile and the locality requirement scenario. Singleton-type minimum distance bounds have been presented and their tightness has been shown by an optimal construction achieving the equality in the bounds. Feasible rate regions have also been characterized by the dimension upper bounds that do not depend on the minimum distance.

## APPENDIX A

### PROOF OF LEMMA 4

*Proof:* We have  $\mathbf{u} = (f(x_1), \dots, f(x_k)) \in \mathbb{F}_{q^t}^k$ ,  $\text{rank}(\{x_1, \dots, x_k\}) = k$ , and  $\mathbf{v} = \mathbf{u}G$ . Without loss of generality, denote the  $s$  symbols in  $\mathbf{v}$  as  $\{v_1, \dots, v_s\}$ , i.e.,  $\mathcal{T} = [s]$ . We get

$$(v_1, \dots, v_s) = \mathbf{u}G|_{[s]} = \left( \sum_{i=1}^k g_{i,1}f(x_i), \dots, \sum_{i=1}^k g_{i,s}f(x_i) \right) \stackrel{(3)}{=} \left( f\left(\sum_{i=1}^k g_{i,1}x_i\right), \dots, f\left(\sum_{i=1}^k g_{i,s}x_i\right) \right).$$

Clearly,  $v_j$  corresponds to an evaluation of  $f(\cdot)$  at  $y_j = \sum_{i=1}^k g_{i,j}x_i$ , for  $j \in [s]$ . Furthermore, we get

$$(y_1, \dots, y_s) = \left( \sum_{i=1}^k g_{i,1}x_i, \dots, \sum_{i=1}^k g_{i,s}x_i \right) = (x_1, \dots, x_k)G|_{[s]},$$

and therefore

$$\text{rank}_{\mathbb{E}}(\mathcal{T}) = \text{rank}(\{y_1, \dots, y_s\}) = \text{rank}(G|_{[s]}) = \min(s, k).$$

■

## APPENDIX B

### PROOF OF LEMMA 7

*Proof:* Given a code satisfying the locality requirement  $(\mathbf{n}, \delta)$  partition  $\mathcal{N}_j$ , for each  $j \in [r^*]$ , into disjoint subsets  $\mathcal{N}_{j,\hat{j}}$ ,  $\hat{j} \in [j]$ , such that for  $i \in \mathcal{N}_{j,\hat{j}}$ , we have  $\text{loc}_{\delta}(i) = \hat{j}$ . Note that  $\hat{\mathcal{N}}_{\hat{j}} = \bigsqcup_{j'=\hat{j}}^{r^*} \mathcal{N}_{j',\hat{j}}$ ,  $\hat{j} \in [r^*]$ . We have

$$\bigsqcup_{\hat{j}=1}^j \hat{\mathcal{N}}_{\hat{j}} = \bigsqcup_{\hat{j}=1}^j \bigsqcup_{j'=\hat{j}}^{r^*} \mathcal{N}_{j',\hat{j}} \supset \bigsqcup_{\hat{j}=1}^j \bigsqcup_{j'=\hat{j}}^j \mathcal{N}_{j',\hat{j}} = \bigsqcup_{j'=1}^j \bigsqcup_{\hat{j}=1}^{j'} \mathcal{N}_{j',\hat{j}} = \bigsqcup_{j'=1}^j \mathcal{N}_{j'},$$

and therefore  $\sum_{j'=1}^j \hat{n}_{j'} = \left| \bigsqcup_{j'=1}^j \hat{\mathcal{N}}_{j'} \right| \geq \left| \bigsqcup_{j'=1}^j \mathcal{N}_{j'} \right| = \sum_{j'=1}^j n_{j'}$ ,  $j \in [r^*]$ .

<sup>3</sup> There is a slight deviation in the boundary conditions, but one can check that this makes no difference and both bounds coincide.

---

**Algorithm 3** Used in the Proof of Lemma 9
 

---

- 1: **while**  $\exists i, j \in [\mathcal{L}], i < j$ , such that  $|\mathcal{R} \cap \mathcal{G}_i| < |\mathcal{G}_i|$  and  $|\mathcal{R} \cap \mathcal{G}_j| > 0$  **do**
  - 2:   Construct  $\Delta\mathcal{R}_i$  and  $\Delta\mathcal{R}_j$  such that
 
$$\Delta\mathcal{R}_i \subset \mathcal{G}_i \setminus \mathcal{R},$$

$$\Delta\mathcal{R}_j \subset \mathcal{R} \cap \mathcal{G}_j, \text{ and}$$

$$|\Delta\mathcal{R}_i| = |\Delta\mathcal{R}_j| = \min(|\mathcal{G}_i \setminus \mathcal{R}|, |\mathcal{R} \cap \mathcal{G}_j|)$$
  - 3:    $\mathcal{R} = \mathcal{R} \sqcup \Delta\mathcal{R}_i \setminus \Delta\mathcal{R}_j$
  - 4: **end while**
- 

Conversely, starting from  $j = 1$ , construct the sets  $\mathcal{N}_j$ ,  $j \in [r^*]$ , by choosing  $n_j$  symbols with  $\text{loc}_\delta(\cdot)$  being at most  $j$ , i.e.,  $n_j$  symbols from the sets  $\hat{\mathcal{N}}_{j'}$  such that  $j' \leq j$ . This is possible since the number of such symbols still left is  $\sum_{j'=1}^j \hat{n}_{j'} - \sum_{j'=1}^{j-1} n_{j'} \geq n_j$ . ■

APPENDIX C  
PROOF OF LEMMA 9

*Proof:* An erasure pattern can be represented by an index set  $\mathcal{R}$  corresponding to the remaining symbols. Any erasure pattern can be transformed into the claimed worst case pattern by invoking Algorithm 3. In Step 3 of the algorithm, as many symbols as possible in the local group  $\mathcal{G}_j$  are replaced with symbols in the local group  $\mathcal{G}_i$ , where  $|\mathcal{G}_i| \leq |\mathcal{G}_j|$ . We claim that that this replacement always yields an erasure pattern with a non-increasing *remaining rank*, i.e., a non-decreasing number of *rank erasures*, making the claimed pattern be the worst indeed.

First, observe that

$$\mathcal{R} = \mathcal{R}_0 \sqcup \mathcal{R}_i \sqcup \mathcal{R}_j,$$

where

$$\mathcal{R}_0 = \mathcal{R} \setminus (\mathcal{G}_i \sqcup \mathcal{G}_j),$$

$$\mathcal{R}_i = \mathcal{R} \cap \mathcal{G}_i,$$

$$\mathcal{R}_j = \mathcal{R} \cap \mathcal{G}_j.$$

We have

$$\text{rank}_E(\mathcal{R}) \stackrel{(a)}{=} \text{rank}_E(\mathcal{R}_0) + \text{rank}_E(\mathcal{R}_i) + \text{rank}_E(\mathcal{R}_j), \quad (35)$$

where (a) is due to Remark 5. Similarly, for

$$\begin{aligned} \mathcal{R}' &\triangleq \mathcal{R} \sqcup \Delta\mathcal{R}_i \setminus \Delta\mathcal{R}_j \\ &= \mathcal{R}_0 \sqcup \mathcal{R}'_i \sqcup \mathcal{R}'_j, \end{aligned}$$

where

$$\begin{aligned}\mathcal{R}'_i &= \mathcal{R}' \cap \mathcal{G}_i = \mathcal{R}_i \sqcup \Delta \mathcal{R}_i, \\ \mathcal{R}'_j &= \mathcal{R}' \cap \mathcal{G}_j = \mathcal{R}_j \setminus \Delta \mathcal{R}_j,\end{aligned}$$

we can write

$$\text{rank}_{\mathbb{E}}(\mathcal{R}') = \text{rank}_{\mathbb{E}}(\mathcal{R}_0) + \text{rank}_{\mathbb{E}}(\mathcal{R}'_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}'_j). \quad (36)$$

From (35) and (36), we have to show that

$$\text{rank}_{\mathbb{E}}(\mathcal{R}_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}_j) \geq \text{rank}_{\mathbb{E}}(\mathcal{R}'_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}'_j). \quad (37)$$

Note that by Lemma 4, we have

$$\text{rank}_{\mathbb{E}}(\mathcal{R}'_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}'_j) = \min(|\mathcal{R}_i| + \Delta, \text{rank}_{\mathbb{E}}(\mathcal{G}_i)) + \min(|\mathcal{R}_j| - \Delta, \text{rank}_{\mathbb{E}}(\mathcal{G}_j)), \quad (38)$$

where  $\Delta = |\Delta \mathcal{R}_i| = |\Delta \mathcal{R}_j| = \min(|\mathcal{G}_i \setminus \mathcal{R}|, |\mathcal{R} \cap \mathcal{G}_j|) = \min(|\mathcal{G}_i| - |\mathcal{R}_i|, |\mathcal{R}_j|)$ .

*Case 1:*  $|\mathcal{R}_i| \leq \text{rank}_{\mathbb{E}}(\mathcal{G}_i)$  and  $|\mathcal{R}_j| \leq \text{rank}_{\mathbb{E}}(\mathcal{G}_j)$ .

By using Lemma 4, we get

$$\text{rank}_{\mathbb{E}}(\mathcal{R}_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}_j) = |\mathcal{R}_i| + |\mathcal{R}_j|.$$

In the case of  $|\mathcal{G}_i| - |\mathcal{R}_i| \geq |\mathcal{R}_j|$ , it is easy to see that (38) leads to

$$\text{rank}_{\mathbb{E}}(\mathcal{R}'_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}'_j) \leq |\mathcal{R}_i| + |\mathcal{R}_j|,$$

which implies (37). Otherwise, we obtain the same result by noting that each term in (38) is upper bounded by  $|\mathcal{G}_i|$  and  $|\mathcal{R}_i| + |\mathcal{R}_j| - |\mathcal{G}_i|$ , respectively.

*Case 2:*  $|\mathcal{R}_i| \leq \text{rank}_{\mathbb{E}}(\mathcal{G}_i)$  and  $|\mathcal{R}_j| > \text{rank}_{\mathbb{E}}(\mathcal{G}_j)$ .

According to Lemma 4, we have

$$\text{rank}_{\mathbb{E}}(\mathcal{R}_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}_j) = |\mathcal{R}_i| + \text{rank}_{\mathbb{E}}(\mathcal{G}_j).$$

From (38), if  $|\mathcal{G}_i| - |\mathcal{R}_i| \geq |\mathcal{R}_j|$ , we get

$$\begin{aligned}\text{rank}_{\mathbb{E}}(\mathcal{R}'_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}'_j) &\leq \text{rank}_{\mathbb{E}}(\mathcal{G}_i) \leq \text{rank}_{\mathbb{E}}(\mathcal{G}_j) \\ &\leq |\mathcal{R}_i| + \text{rank}_{\mathbb{E}}(\mathcal{G}_j),\end{aligned}$$

and therefore (37). Otherwise, we again have

$$\begin{aligned}\text{rank}_{\mathbb{E}}(\mathcal{R}'_i) + \text{rank}_{\mathbb{E}}(\mathcal{R}'_j) &\leq \text{rank}_{\mathbb{E}}(\mathcal{G}_i) + |\mathcal{R}_i| + |\mathcal{R}_j| - |\mathcal{G}_i| \\ &= |\mathcal{R}_i| + |\mathcal{R}_j| - (\delta - 1) \\ &\leq |\mathcal{R}_i| + |\mathcal{G}_j| - (\delta - 1) \\ &= |\mathcal{R}_i| + \text{rank}_{\mathbb{E}}(\mathcal{G}_j).\end{aligned}$$

Case 3:  $|\mathcal{R}_i| > \text{rank}_E(\mathcal{G}_i)$  and  $|\mathcal{R}_j| \leq \text{rank}_E(\mathcal{G}_j)$ .

Lemma 4 gives

$$\text{rank}_E(\mathcal{R}_i) + \text{rank}_E(\mathcal{R}_j) = \text{rank}_E(\mathcal{G}_i) + |\mathcal{R}_j|,$$

and it is immediate from (38) that

$$\text{rank}_E(\mathcal{R}'_i) + \text{rank}_E(\mathcal{R}'_j) \leq \text{rank}_E(\mathcal{G}_i) + |\mathcal{R}_j|,$$

hence (37).

Case 4:  $|\mathcal{R}_i| > \text{rank}_E(\mathcal{G}_i)$  and  $|\mathcal{R}_j| > \text{rank}_E(\mathcal{G}_j)$ .

It is easy to see that

$$\text{rank}_E(\mathcal{R}_i) + \text{rank}_E(\mathcal{R}_j) = \text{rank}_E(\mathcal{G}_i) + \text{rank}_E(\mathcal{G}_j),$$

by Lemma 4, and also that

$$\text{rank}_E(\mathcal{R}'_i) + \text{rank}_E(\mathcal{R}'_j) \leq \text{rank}_E(\mathcal{G}_i) + \text{rank}_E(\mathcal{G}_j),$$

from (38), resulting in (37). ■

## APPENDIX D

### PROOF OF LEMMA 10

*Proof:* Let us denote the evaluation points corresponding to  $\mathcal{T}$  as  $\{y_1, \dots, y_{|\mathcal{T}|}\}$ , where  $y_i \in \mathbb{F}_q^t$  (or equivalently  $y_i \in \mathbb{F}_{q^t}$ ),  $i \in [|\mathcal{T}|]$ . Without loss of generality, we assume that the set  $\{y_1, \dots, y_{\text{rank}_E(\mathcal{T})}\}$  is a basis for the vector space  $\text{span}(\{y_1, \dots, y_{|\mathcal{T}|}\})$ . Then, for  $i = \text{rank}_E(\mathcal{T}) + 1, \dots, |\mathcal{T}|$ , we have

$$y_i = \sum_{j=1}^{\text{rank}_E(\mathcal{T})} \lambda_{ij} y_j,$$

where  $\lambda_{ij} \in \mathbb{F}_q$ . The generator submatrix corresponding to the symbols indexed by  $\mathcal{T}$  can be written as

$$G|_{\mathcal{T}} \triangleq (\mathbf{g}_1^T \ \dots \ \mathbf{g}_{|\mathcal{T}|}^T) = \begin{pmatrix} y_1^{q^0} & \cdots & y_{|\mathcal{T}|}^{q^0} \\ \vdots & \ddots & \vdots \\ y_1^{q^{k-1}} & \cdots & y_{|\mathcal{T}|}^{q^{k-1}} \end{pmatrix}.$$

Furthermore, for  $i = \text{rank}_E(\mathcal{T}) + 1, \dots, |\mathcal{T}|$ , we can write

$$\mathbf{g}_i^T = \begin{pmatrix} (\sum_j \lambda_{ij} y_j)^{q^0} \\ \vdots \\ (\sum_j \lambda_{ij} y_j)^{q^{k-1}} \end{pmatrix} \stackrel{(3)}{=} \begin{pmatrix} \sum_j \lambda_{ij} y_j^{q^0} \\ \vdots \\ \sum_j \lambda_{ij} y_j^{q^{k-1}} \end{pmatrix} = \sum_{j=1}^{\text{rank}_E(\mathcal{T})} \lambda_{ij} \mathbf{g}_j^T,$$

and therefore

$$\text{rank}_G(\mathcal{T}) = \text{rank}_G(\mathcal{T}') = \text{rank}(G|_{\mathcal{T}'}),$$

where  $\mathcal{T}'$  is the symbol index set corresponds to the evaluation points  $\{y_1, \dots, y_{\text{rank}_E(\mathcal{T})}\}$ , i.e.,  $G|_{\mathcal{T}'} = (\mathbf{g}_1^T \dots \mathbf{g}_{\text{rank}_E(\mathcal{T})}^T)$ . Note that  $G|_{\mathcal{T}'}$  is a *Moore matrix* [22], [25] of size  $k \times \text{rank}_E(\mathcal{T})$  with all the elements in the first row being linearly independent over  $\mathbb{F}_q$ . Therefore, an arbitrary square submatrix of  $G|_{\mathcal{T}'}$  is nonsingular, and we get

$$\text{rank}_G(\mathcal{T}) = \min(\text{rank}_E(\mathcal{T}), k),$$

from which, one can easily verify that the claim follows. ■

## REFERENCES

- [1] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, “Xoring elephants: novel erasure codes for big data,” in *Proceedings of the 39th international conference on Very Large Data Bases*, 2013, pp. 325–336.
- [2] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, “On the locality of codeword symbols,” *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6925–6934, Nov 2012.
- [3] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, “Optimal linear codes with a local-error-correction property,” in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 2776–2780.
- [4] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, “Codes with local regeneration and erasure correction,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4637–4660, Aug 2014.
- [5] D. S. Papailiopoulos and A. G. Dimakis, “Locally repairable codes,” *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 5843–5855, Oct 2014.
- [6] I. Tamo and A. Barg, “A family of optimal locally recoverable codes,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4661–4676, Aug 2014.
- [7] S. Kadhe and A. Sprintson, “Codes with unequal locality,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 435–439.
- [8] —, “Codes with unequal locality,” *CoRR*, vol. abs/1601.06153, 2016. [Online]. Available: <http://arxiv.org/abs/1601.06153>
- [9] A. Zeh and E. Yaakobi, “Bounds and constructions of codes with multiple localities,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 640–644.
- [10] —, “Bounds and constructions of codes with multiple localities,” *CoRR*, vol. abs/1601.02763, 2016. [Online]. Available: <http://arxiv.org/abs/1601.02763>
- [11] M. Kuijper and D. Napp, “Erasure codes with simplex locality,” *CoRR*, vol. abs/1403.2779, 2014. [Online]. Available: <http://arxiv.org/abs/1403.2779>
- [12] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, “Optimal locally repairable codes and connections to matroid theory,” in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 1814–1818.
- [13] N. Silberstein, A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath, “Optimal locally repairable codes via rank-metric codes,” in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 1819–1823.
- [14] W. Song, S. H. Dau, C. Yuen, and T. J. Li, “Optimal locally repairable linear codes,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 1019–1036, May 2014.
- [15] S. Goparaju and R. Calderbank, “Binary cyclic codes that are locally repairable,” in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 676–680.
- [16] I. Tamo, A. Barg, S. Goparaju, and R. Calderbank, “Cyclic lrc codes and their subfield subcodes,” in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1262–1266.
- [17] J. Hao, S. T. Xia, and B. Chen, “Some results on optimal locally repairable codes,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 440–444.
- [18] T. Ernvall, T. Westerback, R. Freij-Hollanti, and C. Hollanti, “Constructions and properties of linear locally repairable codes,” *IEEE Transactions on Information Theory*, vol. 62, no. 3, pp. 1129–1143, March 2016.
- [19] A. Pollanen, T. Westerback, R. Freij-Hollanti, and C. Hollanti, “Bounds on the maximal minimum distance of linear locally repairable codes,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1586–1590.

- [20] B. Chen, S.-T. Xia, J. Hao, and F.-W. Fu, “Constructions of optimal cyclic  $(r, \delta)$  locally repairable codes,” *CoRR*, vol. abs/1609.01136, 2016. [Online]. Available: <http://arxiv.org/abs/1609.01136>
- [21] È. M. Gabidulin, “Theory of codes with maximum rank distance,” *Problemy Peredachi Informatsii*, vol. 21, no. 1, pp. 3–16, 1985.
- [22] F. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes*. North-Holland Publishing Company, 1977.
- [23] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, “Optimal locally repairable and secure codes for distributed storage systems,” *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 212–236, Jan 2014.
- [24] W. Song and C. Yuen, “Locally repairable codes with functional repair and multiple erasure tolerance,” *CoRR*, vol. abs/1507.02796, 2015. [Online]. Available: <http://arxiv.org/abs/1507.02796>
- [25] D. Goss, *Basic structures of function field arithmetic*. Berlin New York: Springer, 1998.